DOCUMENT RESUME

ED 354 146 SE 052 972

AUTHOR	Meyer, Linda A.; And Others
TITLE	The Development of Science Knowledge in Kindergarten through Second Grade.
INSTITUTION	Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSR-TR-549
PUB DATE	Mar 92
CONTRACT	NSF-MDR-85-50320; OERI-G0087-C1001-90
NOTE	78p.; For a similar report for grades three and four, see SE 052 973.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Classroom Techniques; Concept Formation; Educational Research; *Family Influence; *Independent Reading; Instructional Materials; *Primary Education; Science Curriculum; *Science Education; *Science Instruction; Science Teachers
IDENTIFIERS	Hands on Science Activities; Science Process Skills

ABSTRACT

Trends for kindergarten through second-grade children were identified from data collected in a longitudinal study of how children develop science concepts. The study involved approximately 325 children from three school districts. A heuristic model of science learning was developed representing children's entering ability, home background, home support, teaching processes, instructional material characteristics, and end-of-year performance. Data were collected in each area, pooled, then submitted to LISREL analyses to produce a structural model for science learning at each grade level. Results showed that children's entering ability in science was the best predictor of their end-of-year science learning overall. The number of hours fathers worked was related to entering ability for kindergarten children, with fathers of higher ability children working more hours. Mothers' education level had by far the highest loadings at all grade levels; mothers' occupation, the lowest. Children's participation in science-related home activities made the greatest contribution to end-of-year performance for all grades, whereas experiences with adults and the number of books and magazines in the home affected only kindergartners' performances. No teaching variable had a significant relationship to end-of-year performance of kindergartners, but teachers' use of sustained feedback after children's incorrect responses was a significant factor in first graders' performance. Time spent in science activities and teachers' uses of science application questions contributed significantly to the performance of second graders. Teachers' coverage of content contributed negatively to end-of-year performance for both first and second graders. (Author)

CENTER FOR THE STUDY OF READING

Technical Report No. 549

THE DEVELOPMENT OF SCIENCE KNOWLEDGE IN KINDERGARTEN THROUGH SECOND GRADE

Linda A. Meyer
James L. Wardrop
C. Nicholas Hastings
University of Illinois at Urbana-Champaign

March 1992

University of Illinois at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center, and in part by the National Science Foundation under Grant No. NSF MDR 85-50320, and grants from the Silver Burdett Ginn and Houghton Mifflin publishing companies. The publication does not necessarily reflect the views of the agencies supporting the research.

Abstract

Trends for kindergarten though second-grade children were identified from data collected in a longitudinal study of how children develop science concepts. The study involved approximately 325 children from three school districts. A heuristic model of science learning was developed representing children's entering ability, home background, home support, teaching processes, instructional material characteristics, and end-of-year performance. Data were collected in each area, pooled, then submitted to LISREL analyses to produce a structural model for science learning at each grade level. Results showed that children's entering ability in science was the best predictor of their end-of-year science learning overall. The number of hours fathers worked was related to entering ability for kindergarten children, with fathers of higher ability children working more hours. Mothers' education level had by far the highest loadings at all grade levels; mothers' occupation, the lowest. Children's participation in science-related home activities made the greatest contribution to end-of-year performance for all grades, whereas experiences with adults and the number of books and magazines in the home affected only kindergartners' performances. No teaching variable had a significant relationship to end-of-year performance of kindergartners, but teachers' use of sustained feedback after children's incorrect responses was a significant factor in first graders' performance. Time spent in science activities and teachers' uses of science application questions contributed significantly to the performance of second graders. Teachers' coverage of content contributed negatively to end-of-year performance for both first and second graders.

THE DEVELOPMENT OF SCIENCE KNOWLEDGE IN KINDERGARTEN THROUGH SECOND GRADE

> Society has never needed individuals with gifts in science and technology (science in action) more than it does as the 20th century comes to an end . . . . Absorbing some of this cascade of scientific information, now said to be doubling every five years or so, is every citizen's duty. Yet beyond that considerable task lies the mission of the few remarkable individuals who have the ability to direct and add to the growth of scientific knowledge. (Fort, 1990, p. 665)

This statement summarizes the position in which the United States currently finds itself. There is unanimous agreement that as a nation, we always have produced and probably will continue to produce a "few remarkable individuals" to be world leaders in science. Yet a recent study of science and math performance of students in five countries found that students in the United States placed last in math and close to last in science performance. In another study, conducted by the International Association for the Evaluation of Educational Achievement, ninth graders from the United States placed 14th (along with Singapore and Thailand) among nations in science performance. Taken together, these studies illustrate that as a nation, the United States is falling farther and farther behind other countries (many of them quite underdeveloped by American standards) in science performance. As a 1989 *Time* article put it: "American science education is a shambles" (Tifft, p. 68).

More than 30 years ago, the launch of Sputnik stunned us into recognizing that the United States had lost supremacy in science to the Soviets. In response, science educators began to argue that the methods used to teach science in the public schools were at fault and that teachers needed to teach science differently. Therefore, millions of dollars were appropriated to develop innovative teaching approaches. Once the approaches were developed, there were serious attempts to implement them--attempts that ultimately failed. Mechling and Oliver (1983) explained this failure as follows: "What was intended to be a joyful discovery for students too often turned out to be a lost sojourn into the abstract and difficult" (p. 43).

We argue that it is time to approach the problem of how to teach science differently. To that end, we have been conducting a longitudinal study of how children are taught and how they learn science concepts in American schools. The study has used what Rosenshine and Furst (1973) describe as "a descriptive, correlational, experimental loop." We first collected descriptive data on children's entering ability, home backgrounds, schooling, home support for science knowledge development, and ability at the end of each academic year. Next, we correlated these many variables. Finally, we have used these data to develop causal models of science instruction and learning with the state-of-the-art LISREL (Joreskög, 1978) procedures that allow us to work with correlational data and develop causal models.

This process sends us first into typical elementary school science classrooms to test children and to observe teachers. The results of these observations are time and frequency data that let us describe the characteristics of science instruction in American schools. We next correlate the variables derived from these descriptions to student outcomes, a variety of student measures that sample what they have been taught as well as general science concepts and procedures. The final step in this sequence is the LISREL models with all of the students participating in the study.

We believe that it is important to document typical science ability for children entering school as well as the instruction they receive. It is equally important to learn about the relationships between each of these variables before we can understand science instruction in American schools today, much less make recommendations for changes.

Our study began when children entered kindergarten. Data collection is ending as the children complete sixth grade. This is the first of two reports presenting results from the first five years of the study, kindergarten through fourth grade. These results are tentative because of the statistics used and because we pooled the descriptive findings from the three school districts participating in the study before calculating the correlations and developing the LISREL models.

It is our belief that this study fills a void in the literature. Most research on elementary science teaching or learning focuses upon students' scientific misconceptions (e.g., Champagne, 1991), properties of science textbooks (e.g., Gilbert, 1989; Strube, 1989), or studies of innovative materials, particularly computer packages, for teaching science (Hofmeister, Engelmann, & Carnine, 1989; Wainwright, 1989). In contrast, this study documents how science is being taught in elementary schools in America today, and the impact that instruction is having upon children's developing science knowledge in comparison to their entering ability, home backgrounds, and home support from their families.

Related Research

The landmark work of Inhelder and Piaget (1958) reported in *The Growth of Logical Thinking from Childhood to Adolescence* and in *The Child's Conception of Space* (Piaget & Inhelder, 1963) provides the research foundation for how children develop reasoning ability. These exceedingly careful reports of natural development are the benchmarks against which we can make important assessments of children's development and against which we can study the impact of differences in children's environments such as home support, classroom instruction, and independent reading. Because children do not develop in a vacuum, each of these forces may be thought of as "acting on" them at any given time.

Research on the consequences of schooling now refutes the often-cited findings of Coleman et al. (1966), who declared that schools could not alter the academic directions of children's lives. In fact, many studies have documented the general effects of schooling as well as specific characteristics of instruction that mediate children's learning. Carroll's (1963) work tying opportunity to learn to content coverage, and the work of several other researchers such as Block (1971) and Anderson (1973) suggests that exposure to schooling influences learning. Work in these experimental settings is supported by the naturalistic classroom observational research of Stallings (1976), for example, who found higher reading performance for first and third graders in classrooms where teaching was characterized by systematic instruction, positive reinforcement, and children's engagement in reading tasks. Similar results have also been found for reading in the work of Brophy and his colleagues both in descriptive (Brophy & Evertson, 1976) and experimental work (Anderson, Evertson, & Brophy, 1978), in math by Good and Grouws (1979), and in both reading and math in the Beginning Teacher Evaluation Study of Fisher et al. (1980). Even more robust findings have come from research looking at instructional time and content coverage, particularly Dahloff's (1971) work in math and the work of Barr and Dreeben (1983) in reading.

Work quite germane to ours on the general effects of schooling was conducted by Heyns (1978) who determined that "schooling has a substantial independent effect on the achievement of children and that outcomes resulting from schooling are far more equal than those that would be expected based on the social class and racial origins of sample children" (p. 9). Heyns conducted a longitudinal study of sixth and seventh graders from Atlanta during the summer, focusing on measuring students' spring and fall achievement. In this way, she contrasted learning when schools were open to learning that occurred when they were closed for the summer. She found that students' summer learning was much more dependent upon family backgrounds than was learning made while schools were in session. Heyns argued that the more equal student growth during the school year supported the notion of schooling effects.

Two other longitudinal studies have addressed the contributions of home and schooling effects on the achievement of young children. Epstein (1980) examined the effects of involvement in school and family life. Measuring children's participation in family decisions, their restrictions at home because of formal rules, their participation in classroom decisions, and their self-direction in classroom instruction, she found that sequential and steadily increasing participation by children in decision making was essential for continued progress in developing mature attitudes and behaviors.

Entwisle and Hayduk (1982) also reported on the cognitive and affective outcomes of early schooling "to discover the extent to which children's achievement in reading and arithmetic in the early grades responds to parents' expectations, peers' expectations, the evaluation process, and/or the cultural context of the school" (p. 23). They found that teachers' expectations influenced children's expectations very little, however, the teachers' "evaluation" component proved very important. Entwisle and Hayduk also found peer popularity to be quite randomly related to children's expectations about school. In addition, the variables they used to determine parents' influences did not prove to be useful. Teachers' influences were limited to their grading practices with the marks they gave for conduct revealing gender differences that replicated previous research findings. Entwisle and Hayduk also found children's expectations persisted only on a limited basis. In summary, they found that children's academic expectations showed only limited persistence from first to third grade.

In 1977, Sorenson and Hallinan used the Project TALENT data bank to reconceptualize schooling effects. Their goal was to develop a model that could ascertain school effects if any existed. Their assumption was that such a model must encompass data from teachers and textbooks. Furthermore, they asserted that teaching is a communication process and that learning can take place only when opportunities for interactions exist. Sorenson and Hallinan set forth their model expecting ability, effort, and opportunity to produce variations in learning. Testing their model with English and math achievement data, however, they received ambiguous results. The model was validated with only the math data. Sorenson and Hallinan argued that "longitudinal data that include appropriate measures of achievement as well as individual level variables and measures of a number of school resources are not available" (p. 285). They argued that those kinds of data are necessary for a fair test of the effects of schooling.

While our study was not designed to meet the criteria specified by Sorenson and Hallinan, it does include the numerous aspects of data collection necessary for a fair test of the effects of entering ability, science textbooks, and classroom instruction, as well as home background and home support for childrens' development of science knowledge. We believe that this study is uniquely able to address these questions and, therefore, makes a substantial contribution to the literature, particularly because we began to collect data when the children entered kindergarten, and, most important, because it is naturalistic, ecologically valid, and data-driven. It includes numerous measures of children's ability; classroom observational data at the child, group, and classroom level on each child in the study; a page-by-page analysis of textbooks; and questionnaire responses from parents that attempt to measure their informal science processes with their children, the science-related activities they engage in, and the books and other science resources they provide for their children. In addition, the longitudinal design allows us to study the continuing process of development in each of these areas, unlike either descriptive or experimental work that is dependent upon data gathered at one or two points in time and then analyzed with statistics more appropriate for a cross-sectional design.

Research Questions

How do children develop science knowledge in kindergarten, first grade, and second grade? As we have ferreted out an answer to this question, several more focused research questions have emerged: What sort of home experiences develop children's factual science knowledge and knowledge of scientific processes? What is the character of these activities? What sorts of things do children choose to do independently that contribute to their acquisition of science knowledge? How much science-related instruction do children receive in the early elementary grades? What are the characteristics of this science instruction? And finally, How do activities at home and activities in school jointly influence the development of children's scientific knowledge?

It is our hypothesis that schooling will explain variance in children's scientific knowledge in the early elementary grades only as a function of the amount of time teachers spend in science instruction while interacting with the children about their background knowledge; giving demonstrations with manipulative materials; asking application questions; having the students formulate hypotheses; and asking questions that center on text that are both explicit and implicit. We predict that the most successful elementary grade science instruction will incorporate hands-on *and* text-based instruction. In the absence of such instruction, student achievement in science will be predicted by the children's entering ability and home background variables.

Heuristic Model of Science Knowledge Development

The simplest way to think about the development of science knowledge is to view it as a function of the aptitudes and abilities children possess as they enter school. That is, what and how much children learn about science in the early grades directly reflects their aptitude or verbal ability. A somewhat more complex view sees learning as a reflection both of children's aptitudes and of home-related factors, such as socioeconomic status and the amount and kind of stimulation provided by parents. Science knowledge development can also be seen as a function of the amount and kind of instructional activities children receive in school. In addition, students may be viewed as contributing to their own knowledge development through the reading they do, the television programs they watch, and the activities they choose.

These separate formulations, when combined, more accurately model the nature of science knowledge development. That is, science knowledge development is a function not only of aptitude and of immediate support and instruction, both from home and school, but of student-initiated activities, such as independent reading.

To address the question of how children develop science knowledge over time, the senior members of our research team developed a heuristic model to guide data collection and analyses at each grade level (see Figure 1).

[Insert Figure 1 about here.]

The model is composed of eight constructs: home background characteristics, student ability at the time they began school, the characteristics of instructional materials used to teach science, teachers' management style, teachers' instructional style, home support for science knowledge development, student ability at the end of each year, and independent reading.

The following discussion explains how we conceptualized each construct.

1. **Home background.** This construct represents the variables of parental occupation and education, the number of adults in the home, the number of older and younger siblings, and the number of hours each parent works outside the home each week.

2. **Ability, 0.** This construct represents children's verbal abilities upon entering school that are most likely to affect their science knowledge at the end of kindergarten. Ability 0 on the model represents the children's abilities at the first testing in the fall.

3. **Materials.** This construct represents the characteristics of instructional materials that may contribute to children's science development. Specifically, it represents the variables of textbook content, use, and "considerateness"--that is, the number of problems in the way science textbooks present material.

4. **Management style.** Because we believed that teachers do not necessarily manage their classrooms in the same ways they instruct, we separated teaching initially into two constructs, management style and instructional style. Management style captures teachers' strategies for molding students' general behavior. It is composed of five classroom characteristics: (a) the amount of time teachers allocated to science instruction; (b) their general praise statements to individual students; (c) their general praise statements to groups of students, such as, "Everyone is working very nicely"; (d) their critical statements directed to individual students, such as "Johnny, sit down and start to work now"; and (e) their critical statements directed to groups of students.

5. **Instructional style.** This is the second construct representing teaching. The variables of the instructional style construct are extensions of characteristics reported in research on general teaching effectiveness in other areas of elementary education such as reading and math that have demonstrated the effects that instruction can produce in these areas. Six additional classroom process variables compose this construct. Half of these variables are measures of the kinds of interactions teachers initiated with individual students or entire classes. The remaining half of these variables capture teachers' responses to students who have made errors or who can not come up with an answer, that is by the kinds of feedback--sustained, terminating, or confirming--that teachers give.

6. **Home support for science knowledge.** This construct contains three clusters of variables: (a) a child's involvement in science processes with parents, (b) the frequency with which parents provide activities for their children, and (c) the prevalence of science-related books and magazines in the home to which the child has direct access.

7. **Independent reading.** This construct represents reading initiated by the child. We anticipated that activities such as independent reading might influence children's science knowledge in later grades, but probably not in the kindergarten through second-grade years.

8. **Ability, 1.** This construct represents students' science knowledge in the spring of each school year.

Method

The Setting

All of the students, parents, and teachers at designated grade levels in three school districts have taken part in this research. In each district, we have studied two cohorts (age groups) of students. Two of the districts had only one elementary school. The third district had 10 elementary schools but only one of these participated in the study. These three districts were selected because of the natural variations they provide in educational philosophies and approaches and instructional materials. We will show how these differences in approaches were reflected in differences in time and other instructional characteristics of science teaching.

District A. District A is in a fairly self-contained small town in the Midwest. Students in the school studied have mixed abilities upon entering kindergarten. There were approximately 90 children per cohort in three classes at most grade levels. This district is known for its high student performance in reading and its average student performance in science in the early elementary grades. The district's educational philosophy includes whole-class instruction in all subjects beginning with an academic kindergarten. Children are almost never grouped for instruction in District A, as district policy dictates that all students at a grade level will cover the same content each year. Regular kindergarten and first-grade classroom teachers in this district maintain primary responsibility for children with special needs in their classes at the lowest elementary grade levels and often gear many of their instructional interactions and feedback to the lower performing students. Merrill's Accent on Science (Sund, Adams, & Hackett, 1982) is the text adopted districtwide.

District B. District B is in a small town near a larger town in the Midwest to which many of its citizens commute to work. The school from this district accounts for about 150 children in Cohort 1. The district has a tradition of average student achievement in reading and better-than-average performance in science in the lower elementary grades. Teachers here begin grouping for some instruction in kindergarten, and this practice continues through all of the elementary grades. Classes are divided into as many as five or six groups for reading instruction in first and second grade. Thus, students in these classes, on average, spend considerable time each day in independent work. Special teachers begin to play important roles with lower performing children in first grade. There is also some tracking of students into self-contained classes even at the first-grade level.

The year Cohort 1 was in second grade saw dramatic changes in District B's science instruction. Prior to that year, the district had no textbook for science instruction in the lower elementary grades. In 1985-86, the district piloted the Holt science program (Abruscato, Fossaceca, Hassard, & Peck, 1986), and a regular classroom teacher was appointed at each grade level to coordinate science instruction. Each coordinating science teacher had the responsibility of collecting materials teachers needed to implement the activities and experiments in the Holt program. This teacher leader also distributed materials to all other teachers at the grade level. The teachers also used their science program's tests on a regular basis to evaluate student performance in science.

District C. The school participating from District C, located in a suburb of a major midwestern city, has many characteristics of an urban school. It is recognized as the highest performing elementary feeder school to its middle school. It has a heterogeneous population, with students from mixed socioeconomic and ethnic backgrounds. White, Black, and Hispanic children made up this portion of the sample. There were approximately 85 children per cohort in three classes that were grouped as combined grade-level homerooms during first and second grade and again during fourth and fifth grade. Kindergartens and third grades were self-contained. Special teachers played a very important role in this school in District C. Bilingual instruction began in kindergarten for children whose parents chose it. These children were pulled out of their regular classes for about 30 minutes each morning and again each afternoon for instruction in English and Spanish. Teachers worked with low stanine children as a separate group from first grade on.

Science was taught in homerooms. For first and second graders, these homerooms spanned two grade levels. Teachers' committees used the Silver Burdett science textbooks (Mallinson, Mallinson, Smallwood, & Valentine, 1985) at each grade level to develop units for instruction in homerooms. Science lessons for first and second graders were taken from the first- and second-grade books. On the days we observed science instruction in kindergarten through second grade, teachers most often concentrated on verbal presentations of science concepts with a strong emphasis on students' background knowledge. Teachers also focused on activities and experiences during these lessons. They seldom referred to science textbooks in either first or second grade, and therefore the textbooks were almost never seen in use in these classrooms.

As these short descriptions reveal, there was substantial natural variation within each of these districts. Each of the districts is fairly typical of numerous school districts in the United States, and together they yield a composite of characteristics that typify American elementary schools today.

Measures

To track the development of children's science concept acquisition, we have administered a number of psychometric measures. While we have given some standardized measures of verbal performance and basic science understanding, we have found that such published tests do not entirely satisfy the needs of this study. Therefore, we developed a number of customized measures of verbal performance and basic science knowledge with algorithms specifically for use within this context. The following description of the standardized measures of verbal performance and basic science knowledge used in the study will be brief because detailed accounts of each can be found in the manuals and reports provided by their publishers. The description of the custom-developed measures will also be brief, because they have been described in some detail elsewhere (Hastings, Meyer, & Linn, 1987; Hastings, Meyer, Linn, & Wardrop, in press). Table 1 presents means and standard deviations for the sample on each measure.

[Insert Table 1 about here.]

Standardized Measures of Verbal Performance

While the primary thrust of this research was to account for variance in children's acquisition and application of science concepts as opposed to their learning of rote scientific facts and vocabulary, we believed that a certain level of verbal competence was necessary for children to understand basic science concepts. Therefore, we included several measures of verbal-reading performance in our model. Several of th se were standardized tests that have been nationally normed.

Wide Range Achievement Test. We administered the reading subtest from Level I of the Wide Range Achievement Test (WRAT) at least once a year during the course of this study. Cohort 1 was given the 1978 version (Jastak, Jastak, & Bijou) in kindergarten and first grade, and the 1984 version (Jastak & Wilkinson) beginning in second grade. Items on the WRAT consist of a series of increasingly difficult words that children read aloud to the examiner. The measure is individually administered and has a stopping rule whereby 12 consecutive errors terminate administration. Although the items are the same for both editions of this test, the norms differ. Fall WRAT scores were used as manifest variables of the latent variable beginning verbal performance at each grade level.

Reading Mastery Tests

Woodcock. The reading comprehension passage subtest of the Woodcock Reading Mastery Tests (Woodcock, 1973) has been administered numerous times to Cohort 1 children in the longitudinal study. The basic format of this subtest consists of increasingly difficult cloze passages children read. A stopping rule based on 5 consecutive errors is used to terminate administration. Like the WRAT, the Woodcock was used as an observable measure of the latent variable beginning verbal performance at each grade level.

CIRCUS Reading Test. The CIRCUS Reading Test, Level D (Educational Testing Service, 1976a) was given to Cohort 1 children in the spring of their second grade year. This is a relatively traditional group-administered reading test, consisting of a series of short passages followed by comprehension questions.

Customized Measures of Verbal Performance

Although we used standardized tests of verbal performance, we felt that such tests did not measure precisely the latent traits we wished to study. We therefore created one instrument of verbal performance and used two other measures developed by other researchers.

Analogies Subtest of the Language and Problem-Solving Battery. Mason and Meyer (1983) developed the Language and Problem-Solving Battery specifically for the longitudinal study of reading comprehension that runs in tandem with this study. The Analogies subtest of that battery was used as a measure of the verbal ability of children when they entered kindergarten.

Chicago Reading Test. The Chicago Reading Test was developed by Barr (1983). This instrument was given during the winter of Cohort 1's kindergarten and first-grade years and serves as a measure of first-grade verbal performance. It is used to test individually on consonant sounds, vowels and word endings, word families (e.g., *cat*, *fat*, . . . *mat*) and nonsense words (e.g., *fon*).

Interactive Reading Assessment System. The Interactive Reading Assessment System (IRAS) (Calfee & Calfee, 1982) asks students to read up to 14 word lists of 8 words each until the stopping rule based upon accuracy is applied, or the last list is read. Rate, accuracy, and self-corrections are recorded. Students then read up to 7 passages of increasing length and difficulty until they make more than 10 decoding errors and miss half or more of the comprehension questions for a passage. Rate, accuracy, self-corrections, and correctness of responses with or without a prompt to questions based on the passages are recorded for this section also. This measure was used for LISREL modeling as an indicator of beginning of second-grade verbal performance.

Standardized Measures of Basic Science Knowledge

As with the measures of verbal performance, we wanted to use some standardized tests as measures of basic science understanding. A more detailed discussion of the rational for the instruments chosen and the results of their administrations may be found in Hastings et al. (1987) and Hastings, Meyer, Linn, and Wardrop (in press).

Test of Basic Experiences - 2. The Test of Basic Experiences (TOBE-2) (Moss 1978a, 1978b) asks children to choose one of four line drawings in response to orally administered item stems. Level K and Level L of the test were used as end-of-year dependent variables for kindergarten and first grade.

Customized Measures of Basic Science Knowledge

We developed three instruments to test students on content domains from their science textbooks. The textbooks were either at or above grade level. We tried to use vocabulary common to all three school districts and we included out-of-level items in each battery so that we could observe children developing knowledge on a few select science topics. All three measures were initially used as end-of-second-grade manifest variables for basic science understanding.

Motion Test. The Motion Test (Meyer, Hastings, & Linn, 1986) is administered orally and is composed of items exclusively above grade level for second graders. It includes items with balls and mirrors and other objects. The children are asked to select the paths these objects would take to or from a flat surface.

Plants Test. The Plants Test (Meyer, Hastings, Greer, & Linn, 1986a) is orally administered, in that even when items include words, they are read aloud to the children. This instrument contains both in-level and out-of-level items, and common vocabulary from all three school districts. It relies heavily on line drawings.

Three Forms of Matter Test. The Three Forms of Matter Test (Meyer, Hastings, Greer, & Linn, 1986b) also contained in-level and out-of-level items and vocabulary common to the districts, and it made use of line drawings wherever possible.

Measures of Verbal Performance and Basic Science Knowledge

Of the various tests we gave, these measures of verbal performance and of science knowledge might be considered galluses in that they seem to be suspended somewhere between verbal and/or cognitive performance and measures of basic science understanding. Three are nationally normed measures, and one is a measure that we developed specifically for this study.

CIRCUS--Think It Through. The CIRCUS--Think it Through (Educational Testing Service, 1976b) was used as a beginning-of-year variable for both first and second grade. It is a group-administered measure of problem solving. In first grade it loaded only on verbal performance; however, in second grade it loaded on both verbal performance and basic science understanding.

CIRCUS--Listen to the Story. The CIRCUS--Listen to the Story (Educational Testing Service, 1976c) was administered during kindergarten, first grade, and second grade. This is a group-administered test of listening comprehension. In kindergarten, it loaded onto beginning verbal performance. However, in first and second grade it became part of the latent variable basic science knowledge.

Error Detection Test. The Error Detection Test (Meyer, Hastings, Greer, & Linn, 1985) attempts to measure a cognitive domain (detection of errors in written sentences and sequences), number of decoding errors, and children's ability to provide support for their identified errors. The instrument uses vocabulary common to the three school districts in the setting for this study, and depends on science content about plants common to all three districts. It was administered to both first and second graders in the study.

Sequential Test of Educational Progress. The science subtest of the Sequential Test of Educational Progress (Educational Testing Service, 1979) was given to our Cohort 1 children in the spring of their second-grade year. This test is similar in design to the TOBE-2 because it is group- administered and is composed primarily of pictures and small amounts of text. The Step Science subtest tends to load more with "reading measures" than it does with certain "science concept" measures, specifically to TOBE-2, and Motion Test, Plants Test, and Three Forms of Matter Test. This was true for both a promax rotated forced two-factor analysis (Hastings et al., 1987) and a confirmatory two-factor analysis using LISREL (Hastings et al., in press). This may be due to the fact that students must read items on this instrument silently in order to respond.

Procedures

Home background. Information about parents' occupations and education levels and about the family constellation was gathered through questionnaires completed by parents each year of the study.

Ability, 0. We selected students' entering abilities in analogical reasoning and listening comprehension to represent their general verbal ability during the first few weeks of kindergarten. We did so in the belief that these abilities were most likely to affect students' science knowledge at the end of kindergarten.

We assessed analogical reasoning with the Language and Problem-Solving Battery. The analogical reasoning items on this measure fall into three categories: Those that ask students to complete a statement that describes a part-whole relationship ("A person has eyes, a house has _______"), those that children complete on the basis of identifying categories ("A drum is a musical instrument, a necklace is ________"), and those that require children to identify a characteristic ("An inch is short, a mile is ________").

Listening comprehension was assessed with the CIRCUS--Listen to the Story test administered in October of the kindergarten year. About 60% of the items on this test were labeled as "comprehension" measures, and the remaining items were labeled as "interpretation" measures. A typical comprehension item is shown in Figure 2.

[Insert Figure 2 about here.]

A typical interpretation item appears as Figure 3. Total scores from these two instruments were used to compose the latent trait entering verbal ability.

[Insert Figure 3 about here.]

Materials. What characteristics of instructional materials may contribute to children's science development? We chose to study three dimensions of textbooks: "contents," because of previous research on the relationship between content coverage and student performance; how the books are used or "style," because of research supporting activity-based science instruction (Bredderman, 1983) in contrast to text-based programs; and "considerateness" (the number of problems in the way textbook material is presented), because of the numerous reports asserting problems with the quality of science textbooks. An additional dimension of materials we measured was textbook utilization, the number of content domains actually observed as covered during science instruction. The category of utilization may prove to be a particularly important aspect of science materials in the very early elementary grades where teachers pick and choose science topics instead of working through an entire textbook as they are more likely to do in later grades.

"Contents" was measured by counting the number of content domains presented and the number of vocabulary words presented in each science textbook series at each grade level.

"Style" was measured by counting the number of teacher-directed and optional activities as well as the frequencies of background knowledge, text-tied, or review questions provided for teachers to ask. Background knowledge questions are those children answer from information in their heads, text-tied questions are answered in the text, and review questions simply refer back to content presented in previous lessons.

"Considerateness" was measured for each series by applying the Anderson and Armbruster (1984) categories for measuring inconsiderate text. Generally, the process of measuring a text's inconsiderateness involved searching the textbooks for problems such as incorrect information, unclear or unnecessary pictures, or unnecessary figurative language. A complete report of the science textbook analyses is available in Meyer, Crummey, and Greer (1988).

Management style. Do the ways teachers manage their classes affect students' performance in science? Management style was composed of five classroom characteristics. It encompassed the amount of time teachers allocated to science instruction, their general praise statements to individuals or groups of students such as "Everyone is working very nicely," and critical statements directed either to individuals such as "Johnny, sit down and start to work right now," or to groups of students. Additional measures of classroom management style included the number of 5-minute segments teachers were not directing whole classes, and the percentage of students on task during these times.

Instructional style. Our observational system was patterned after procedures developed by others to capture teachers' instructional interactions with each student. Whole days were observed, and time as well as interactions during every part of the school day that were related to science instruction were recorded as observers made sequential scripts of the lessons. This observation system differs from procedures used elsewhere because all children in each classroom were observed and teachers' interactions were coded at the level from which they were delivered. In other words, interactions to individuals were credited to the specific child whom the teacher addressed, questions to an entire class were credited to each child in the class. We also recorded the kinds of interactions teachers initiated with individual students or entire classes, and teachers' responses to students who had made errors or who could not come up with an answer. Teacher-directed interactions were coded as procedural if they gave students directions ("Open your books to page 41."). Text-tied interactions were questions those students answered directly from text they had read or text the teacher had just presented to them. If, for example, a kindergarten teacher said, "One part of a flower is the stem. What is one part of a flower called?", this was coded as a text-tied interaction. If, on the other hand, the teacher said, "Raise your hand if you can tell me the names of three kinds of flowers," this interaction was coded as background knowledge (scriptal), because children must have been able to answer this question from information already stored in their heads.

Instructional style was also characterized by the kinds of feedback teachers gave. Sustained feedback was coded when the teacher persisted with a child or group of children who gave an incorrect response or no response at all until a correct response came forth. To sustain feedback, teachers gave hints, asked for explanations, or led students through a process to come up with a correct answer. Terminating feedback, on the other hand, was what a teacher gave when calling on another child, ignoring an incorrect response, or giving a correct answer in response to the child's incorrect answer. Confirming feedback occurred when a teacher repeated a response or simply said, "Yes, aha."

We observed each kindergarten, first-, and second-grade teacher for 9 full days, scheduling observations from mid-September through mid-April at approximate 3-week intervals. Time spent in science instruction and each interaction during that time was coded at either the individual or class level, depending upon how the teacher taught. Feedback teachers gave was coded in the same way the instructional interactions were coded.

Home support. We used a questionnaire to gather data from parents about home activities that might contribute to children's acquiring basic science understandings. We asked parents to complete the questionnaire for each year of the study. The response rate was never lower than 83% and it actually rose each year the study continued.

The specific indicators of home support activities we used differed somewhat from one year to the next. At the kindergarten level, three indices were used: participation in science-related activities, experiences with adults, and number of science-related books and magazines available. The first two of these indices were carried on through first and second grades, but the number of books and magazines was replaced by an index of literacy-related experiences in first grade, and neither of these was used at the second-grade level.

Beginning in first grade and continuing in second grade, an index of school activities brought home was used to indicate the extent to which students shared with their families the kinds of things they were doing in school.

Finally, for first grade only, we used an index of process/activity experiences. This index was based upon questions parents answered in a series of paragraphs describing short scenarios that might have taken place at home in their kitchens on a Saturday. These scenarios provided settings common to all families. They attempted to capture the amount and type of informal instruction that could take place in a home. A brief portion of the paragraphs in the kindergarten parents questionnaires appears as Figure 4.

[Insert Figure 4 about here.]

The activities' measure was derived from frequencies parents provided for the number of times they took their child to a zoo, a park, a farm and elsewhere.

Ability, 1. Various measures were used to assess science knowledge in the spring of each school year. In kindergarten and first grade, science ability was measured on the group-administered TOBE-2, Levels K and L. In second grade, two content-domain tests, Plants and Three Forms of Matter were used to test children's development of science knowledge in areas they had covered in school. Motion, a third content-domain test, covered content that had not yet been presented in school.

Descriptive Results

Home Background

The mothers and fathers of children participating in this study were likely to have junior college educations, although we found a substantial range in parental background in each of the three districts. District C parents were the best educated. They often had some coursework beyond college. The fathers tended to be businessmen in managerial positions. Mothers of children in Districts A and B tended to be housewives. District C mothers more often worked outside the home.

The children in this study averaged one older brother or sister and half a younger sibling per family. Most households had two adults, although the average was higher than two adults in District C.

Materials

Kindergarten. None of the three school districts had a kindergarten-level textbook for science instruction. Because there were no textbooks, "coverage" information for this level came from the actual classroom observations. The science topics observed were: animal homes in District A; how food grows and what to do when a tornado threatens in District B; and daily weather, information on owls and other birds, raccoons, and animal shelters during the rain in District C.

First grade. Two of the three districts (A and C) had first-grade level textbooks. District A students had the Merrill Accent on Science series, and first graders in District C had the Silver Burdett Science series. Analyses of differences between these textbooks (Meyer, Greer, & Crummey, 1986; Meyer et al., 1988) illustrate dramatic differences between the contents and style of these materials and their similarities with regard to considerateness when applying the Anderson and Armbruster (1984) criteria for considerate text. For example, the Silver Burdett series has 8 content domains and 241 vocabulary words, whereas the Merrill series has 4 content domains and 56 vocabulary words. Similar differences were found between programs for the various activities and types of questions teachers are to ask.

Coverage data from classroom observations revealed that plants, animals, seasons, the water cycle, and lessons on nutrition and care of teeth comprised the topics taught during classroom observations in District A. District B teachers covered primarily these same topics. In addition, they taught lessons on the skeleton. District C teachers focused primarily on plants and animals as well, but they were also observed teaching lessons on colors and shapes, the solar system, and safety.

While science instruction in first grade increased dramatically over science instruction in kindergarten, it was still quite sparse. Several of the lessons observed were actually health rather than science. Still other lessons involved collecting natural materials out-of-doors and then bringing them into the classroom to use in art lessons. One of the initial science lessons observed, in fact, was of this type. A class of children went outside with their teacher in the late fall and collected leaves. When they returned to their classroom, their teacher instructed them to take out their science books. At this point the observer was poised and eager for a science lesson. Then the teacher continued, "Now take out your social studies books." The "lesson" concluded: "Now put your leaves between your science books and your social studies books."

Second grade. All three districts used textbooks to teach science in second grade. District B used the Holt program and appointed a second-grade teacher to oversee science teaching for that grade level. That teacher was also to prepare and collect materials for science activities and experiments to guarantee that all second-grade teachers would be able to teach science regularly. This attention to and preparation for science teaching was unique for the three districts in this study.

The differences between the Silver Burdett and Merrill programs found at the first-grade level carried over into the second-grade materials. The Silver Burdett series covered 10 content domains and emphasized 277 vocabulary words, whereas the Merrill series presented just 9 content domains but emphasized only 44 vocabulary words. The Holt series, in contrast, covered 12 content domains and taught 88 vocabulary words.

Content coverage included lessons on sounds, safety, magnets, animals, and health in District A. District B teachers presented lessons on sound, health, safety, the ocean, insects, animals, evaporation, force, and the skeleton. District C teachers taught lessons on plants, the earth, air and water, health, and animals. Thus, the observational coverage data show overlap for the districts for health and animals. Little observed common content is not surprising given that the textbook analyses revealed few common content domains among publishers at each grade level.

The general findings from this study are that science textbooks at the elementary level vary substantially in the amount and type of content they present as well as in the ways teachers are directed to present the information. An unexpected finding was that the Merrill, Silver Burdett, and Holt textbooks are very considerate. In fact we could find no examples of inconsiderateness in over 40% of the Anderson and Armbruster categories.

Instructional Style

Kindergarten. Results for the three districts for time spent in science instruction in kindergarten appear in Table 2.

[Insert Table 2 about here.]

Time is reported in Table 2 as the average number of minutes of observed science instruction for each of the 9 full-day observations by district. Not only is there substantial variance between districts in the amount of time teachers spent teaching science, but the large standard deviations for districts B and C show substantial variance within districts as well. The great variance within districts is the result of some teachers in each district simply teaching no science at all at the kindergarten level. As the column on the far right of the table indicates, fully half of the District A classes, over 40% of the District B classes, and a third of the District C classes had no observed minutes of science instruction. District C students averaged a little over 2 minutes per observed day in science, while District A students averaged close to 1 minute of science instruction daily, and District B students had only about 30 seconds of science instruction each day observed.

In addition to measuring the number of minutes observed during science instruction in kindergarten, we also wanted to investigate the kind of interactions and feedback teachers had with students during these periods, in other words, what teachers *did* with their time in science. Table 3 presents these results.

[Insert Table 3 about here.]

All of these results are reported at the student level as was the measure of time. Students in District A had the highest number of procedural interactions with their teachers, just under 1 interaction each. Teachers' instructions to children, such as "Open the glass jar," were coded as procedurals. District C children averaged the highest background knowledge and text-tied interactions, also close to 1 per child. Background knowledge (scriptal) interactions are those for which children must come up with the answers from information in their heads. Text-tied interactions, in contrast, are those for which children must search the text to be able to answer the question. Feedback of all types was low in all three districts, generally one fifth of a feedback interaction or less per child. For these analyses, teachers' feedback was divided into three types. *Sustaining feedback* occurred when a teacher did something such as lead a child through a procedure until the child gave the correct answer. *Terminal feedback* stopped interactions between a teacher and a child. *Confirming feedback* was feedback teachers gave that simply repeated what a child said as if to confirm it.

It is important to note that within-district variance is substantial for all types of interactions and feedback. The standard deviations are larger than the means in 16 of the 18 categories in Table 3, indicating substantial positive skewness for these variables.

First grade. Results for the time spent in science instruction in first-grade classes appear in Table 4.

[Insert Table 4 about here.]

The pattern of time spent in first-grade science instruction matched that of the kindergarten classes in the three districts. Once again, District C teachers spent the most time teaching science. They averaged over 8 minutes a day during each observation. District A teachers averaged over 6.5 minutes per observed day, and District B teachers allocated the least amount of time, just over 5.5 minutes daily to science instruction. The standard deviations for time spent in science instruction followed a pattern identical to the mean time for science instruction. District C teachers varied the most, and District B teachers varied least in the amount of time they spent teaching science. In addition, we found that all first-grade teachers in all three districts spent some time teaching science, although there was a substantial range in time from teacher to teacher. Teachers in District A ranged from 12 to 107 total minutes of science instruction for the 9 observation rounds, teachers in District B ranged from 27 to 86 total minutes, and District C teachers taught science from 42 to 140 minutes during observations.

Table 5 presents means and standard deviations for interactions and feedback categories. For these classroom process variables, we see that teachers in District C had both the greatest number of procedural and text-tied interactions, whereas District B teachers asked more background knowledge questions. District B teachers also gave the most sustaining, terminating, and confirming feedback. These same teachers repeated their students' answers just under 3 times (2.71) per child per observation.

[Insert Table 5 about here.]

Thus, at the first-grade level, the picture of science instruction in these three schools shifted substantially from district to district. District C teachers, who allocated the most time to science instruction also had the greatest number of procedural and text-tied interactions but, overall, gave little feedback to students' responses. District B teachers, on the other hand, presented quite a different profile. They spent less time teaching science, but they probed their students' background knowledge more and then provided more support in the form of feedback to these students' responses than did teachers in the other districts.

Second grade. Results by district for time spent in science instruction in second grade appear in Table 6.

[Insert Table 6 about here.]

There are major differences between districts for the amount of time spent in science instruction at the second-grade level. In comparison to the first-grade results, there was a change in the rankings for Districts A and B. District A second-grade teachers actually spent less time in science instruction than did their first-grade colleagues, and their standard deviation for time was close to their mean. In fact, one District A teacher had 0 minutes of observed science instruction over all observations. District B second-grade teachers spent almost twice the number of minutes in science instruction as their first-grade peers. District C teachers increased their time spent in science substantially over that spent by the first-grade teachers, and they continued to spend the most time teaching science of the three districts studied.

Table 7 presents the means and standard deviations by district for interaction and feedback classroom process variables during second-grade science instruction. These data again reveal substantial differences between districts. With the exception of the background knowledge interaction category, District C teachers had more interactions with children than did the teachers in the other districts. A similar pattern was found with feedback. District C teachers gave more sustaining and terminal feedback than did teachers in the other two districts, although District B teachers gave by far the greatest amount of confirming feedback.

[Insert Table 7 about here.]

Management Style

Kindergarten. We based a teacher's general management style on the number of times she or he praised or criticized groups or individual students and on the amount of time the teacher instructed less than the whole class. These aspects of a teacher's management style contributed to the percentage of students working on task when the teacher was not instructing the entire class. At the kindergarten level, one teacher had a very high criticism rate, one that exceeded 600 critical statements to groups and individuals during a single day's observation. The variance in teachers' praising was almost equally as great. Therefore, for these reasons the raw data for praising and criticizing statements were transformed by taking their square roots to restrict the range of frequencies within these categories. After this transformation, praise statements to individual students averaged over 1.5 per child. Praise statements to a class or group averaged just under 1 per child. Individual criticism statements were almost 3.5 per child, and group or class criticism statements averaged about 2.75 to each child. These teachers spent about 35 minutes a day working with less than their whole classes, and the children averaged over 80% (81.12%) of their time on task while working independently.

First grade. Overall management patterns for first-grade teachers were similar to those found for kindergarten teachers. First-grade teachers praised far less than they criticized, however. They averaged about 2.5 individual praise statements and just over 2 (2.14) class praise statements per observation round. On the other hand, their mean number of individual critical statements averaged 48.88 to individuals and just over 27 to classes per round. The teachers averaged over 12 5-minute periods of non-whole class instruction each day they were observed. This means children worked independently on the average just over 1 hour a day. This amount of independent work was most often a function of teachers having reading groups and assigning independent work to all students who were not in the group under the teacher's direction. Children averaged being on task over 87% of the time during their independent work periods.

Second grade. Second-grade teachers' management styles present some interesting contrasts to those of their kindergarten and first-grade colleagues. Their individual praise statements averaged about 2.5 to individuals and 1.6 to groups per day observed. Individual criticisms averaged over 7 and class criticisms (directed toward groups) averaged over 5. Second graders spent 86 minutes each day working independently, and they were on task over 83% of this time.

Home Support

The index of school activities brought home, which was used to indicate the extent to which students shared with their families the kinds of things they were doing in school, showed a gradual increase in school activities taken home by children in first grade (5.87) to second grade (6.88). This increase paralleled the gradual and consistent increase in science instruction observed at these grade levels.

Descriptive results for these indices are summarized in Table 8. The index of participation in activities, which yields scores that range from 0 to 30, had its largest overall mean (18.00) at the kindergarten level, with lower values at first and second grades. In general, there were only small and inconsistent differences among schools on any of these home-support indices.

[Insert Table 8 about here.]

Structural Modeling of Science Knowledge Attainment

In our attempts to account for students' science knowledge attainment at the end of each school year, we employed structural equation modeling using Joreskög's (1978) Linear Structural Relations (LISREL) model. Analyses were performed using the LISREL VI computer package (Joreskög & Sorbom, 1984). The following paragraphs provide a brief overview of the LISREL approach. For a more detailed discussion of the approach, see Joreskög and Sorbom.

The LISREL model incorporates two components: a structural equations model and a measurement model. The structural equations model, like the more traditional path analysis approach, seeks to estimate structural (path) coefficients for a hypothesized model of the relationships among constructs or variables; the measurement model uses confirmatory factor analysis techniques to accommodate the use of multiple indicators for the latent variables (constructs) of interest and to take into account measurement errors and specific variances (in the factor analytic sense) of the observed variables.

Although several statistical methods are available for estimating parameters in the model, we used the maximum likelihood estimation procedure which provides a chi-square index of goodness of model fit, t-statistics for the significance of individual parameter estimates, and modification indices suggesting how much improvement in fit could be obtained by relaxing the various constraints (see below) specified in the current version of a model.

In implementing LISREL modeling, we used a combination of criteria for deciding when a model represented an adequate "fit" to the observed relationships in the data. We sought a ratio of the chi-square statistic to its degrees of freedom of 2.5 or less; a goodness-of-fit index of .90 or more; and a root mean square residual of .06 or less. These criteria are all somewhat arbitrary, but this particular combination seems to represent a reasonable standard that avoids excessive "overfitting" to the unique characteristics of the sample and yet results in models that are theoretically and aesthetically satisfactory.

Within the LISREL approach, there are three kinds of model parameters: free parameters, which are to be estimated from the data; constrained parameters, in which two or more estimates must be equal to each other; and fixed parameters whose values are prespecified. In the analyses reported here, structural parameters were either free (estimated based on observed relationships among variables) or fixed equal to 0 (representing the absence of a path between two variables). For the analyses reported here, measurement model parameters were free when coefficients corresponding to factor loadings were to be estimated from the observed correlations; constrained when two observed variables were required to load equally on a latent variable; fixed equal to 0 when some observed variables were required not to be related to particular latent variables; or fixed equal to 1, when a single indicator was used. In one instance (end-of-kindergarten science performance as measured by the TOBE-2), we fixed a measurement parameter at a value between 0 and 1 (.889), equal to the square root of the estimated reliability of that measure.

As we proceeded through the series of analyses at each grade level, we explored several alternative approaches to specifying the relationships among the classroom variables (invariably the most difficult components to model because the pattern of their intercorrelations with one another was usually inconsistent with their pattern of intercorrelations with the test data used to represent outcomes). We have not included details of these intermediate analyses/models.

The following description of the structural modeling analyses consists of three parts. The first two parts are repeated at each grade level: an initial model, representing an instantiation of the general heuristic model, is described; then the major modifications made during the analysis are summarized, along with presentation of the "final" model that resulted. The third part is a discussion of the trends observed as we moved from kindergarten through first to second grade, with respect to the importance of the "teaching" variables vis-à-vis variables characterizing home and student influences on end-of-year attainment. Before considering the results of our modeling efforts, though, it will be useful to review some of the considerations that guided our work.

General Concerns Relating to Characteristics of the Data

At each grade level, we began with the entire set of data that had been collected. Obviously, in a study of this magnitude, some missing data are inevitable, and in some instances blocks of data for groups of students were missing. (For example, one kindergarten teacher removed the identification labels from students' home questionnaires so that those responses could never be associated with individuals in the sample.) To maintain a sufficiently large sample size, we chose to use all available data in these analyses, so that there were differences from variable to variable in how many and which students provided the data. For this reason, all the LISREL analyses described below use correlation matrices instead of the covariance matrices that are ordinarily preferred.

Univariate Distributions

For some variables, examination of the univariate and bivariate distributions revealed that the data were badly skewed in the positive direction, so a transformation (logarithmic or square root) was needed. This characteristic was consistently observed with the variables indexing teachers' use of praise and criticism at the kindergarten and first-grade levels.

Multi-Collinearity

In addition, because of the substantial multi-collinearity among some variables (especially those involving teachers' classroom behaviors), the initial correlation matrices were invariably near-singular. For example, the determinant of the correlation matrix for the second grade was approximately 0.2×10^{-12}. At both the kindergarten and first-grade levels, it was easy to identify pairs or clusters of variables whose intercorrelations were substantially responsible for this multi-collinearity and to create simple summed composites that were meaningful. Using these composites instead of the separate highly correlated variables partially solved the collinearity problem. At the second-grade level, only teacher praise (to an entire class and to individuals) and teacher criticism (to an entire class and to individuals) could be treated this way, but creating these composites did not correct the problem. (The final second-grade analysis was based on a correlation matrix whose determinant was about 0.2×10^{-9}.) For these data, the only successful strategy was to use extreme care in the specification of initial estimates for the parameters to be estimated.

Kindergarten Analyses

At the kindergarten level, we began with the model portrayed in Figure 5. Major components of the model are:

1. Home background variables of parental education and occupation, numbers of younger and older siblings and adults in the household, and hours worked per week by each parent.

2. Pupil entry-level competencies as measured by the Analogies Test, WRAT, and the listening subtest from the CIRCUS battery.

3. Home-support activities potentially related to the acquisition of science-related understanding: participation in activities such as trips to the zoo, availability of reading materials dealing with science, and involvement with one or both parents in activities around the home that could provide the opportunity for acquiring science-related knowledge (e.g., cooking, gardening).

and several types of teaching variables:

4. Science instructional time.

5. Teacher management style, as reflected in the frequency of praise and criticism directed to both individual students and to the entire class.

6. Teacher instructional style, as represented both in certain types of feedback and interactions and in the frequency with which students were engaged in seatwork.

[Insert Figure 5 about here.]

All of these categories of variables were seen as potentially contributing to students' end-of-year performance on the TOBE-2, used as a measure of understanding of elementary science concepts. The complete set of variables originally considered for the kindergarten analysis is summarized in Table 9, along with their disposition when we arrived at our final model.

[Insert Table 9 about here.]

Home Background Variables

Several of the home background variables we thought might have an effect on end-of-year performance did not. Specifically, those variables indicating numbers of siblings and adults in the home, and the number of hours per week these students' mothers worked turned out to be unrelated to other characteristics in the model. Interestingly, the number of hours per week fathers worked was significantly related to students' beginning-of-kindergarten ability, with fathers of higher ability students working more hours per week. The composite home background variable was more heavily determined by parents' education level than by their occupations, with mother's education receiving the largest weight and the prestige of mother's occupation receiving the smallest.

[Insert Figure 6 about here.]

Home Support

The three observed variables in the home support group loaded very similarly on the composite (latent) variable, with loadings ranging from .52 to .65 (Figure 6). Through its effect on entry-level ability (ß = .29, Figure 6), home support has a total effect on end-of-year performance of .36 (Table 10).

Beginning-of-Year Aptitude

As expected, beginning ability has far and away the strongest influence on final performance. The three tests all contribute quite similarly to definition of the composite pre-kindergarten ability latent variable (loadings ranging from .67 for the CIRCUS listening measure to .74 for Analogies). The structural coefficient from beginning ability to final performance (ß = .81) is indeed substantial, and the total effect of beginning ability approaches 1.00. (See the discussion of effects in the Structural Model section below.)

Classroom Variables

We began our analyses with very modest expectations about the role of the classroom variables, realizing that on the average these kindergartners were exposed to less than 10 minutes per week of what could, generously, be called science-related instruction. Nevertheless, we felt that our initial model should reflect as much information as possible about factors in these children's lives that might contribute to their science knowledge.

In fact, none of the classroom variables had a significant relationship to end-of-year performance, and most relationships of these variables to those in any other category were essentially 0. (We should also note that several of these classroom variables were so highly interrelated that the only feasible way to incorporate them into the modeling was to create a simple summed composite: what we provisionally called "content-oriented behavior" is the sum of background knowledge interactions, text-tied interactions, and science instructional time.) Although their relationships with final performance in science were not significant, most of these variables were retained in the model both as a reminder of the richness of the data available and, more important, because relationships among variables within this category were interesting in their own right.

This structure relating teachers' behaviors and decisions to children's task-oriented behavior is certainly a plausible one. What is lacking, though, is any significant connection of these variables to our measure of end-of-year science understanding the TOBE-2. The small negative relationship between instructional style and final performance (ß = -.07) is not statistically different from 0 (t = -1.60, df = 316). Such relationships simply do not exist for these kindergarten students. (The largest simple correlation of any classroom variable with the TOBE-2 is -.075 for procedural interactions.)

End-of-Year Science Performance

At the kindergarten level, we had only one measure of final performance in science, the TOBE-2. In a separate analysis, we obtained an internal consistency estimate of reliability for this test of .797, so we used the square root of this value (.893) as the fixed coefficient from the observed variable to a latent (true-score) indicator of performance. Subsequent results suggested that this was an underestimate of the test reliability. (See the discussion of Effects in the next section.)

Structural Model

Structural Equation Coefficients. Our final model of the influences on end-of-kindergarten science performance is presented in Figure 6. It is clear from this figure that far and away the major predictor of end-of-year performance is entry-level ability (path coefficient = .813), with a minor *negative* contribution from instructional style (-.073). In this model, the latent variable we have called "instructional style" is a linear combination of--in order of importance--content-oriented behavior, procedural interactions, sustaining feedback, terminating feedback, and confirming feedback. For the model depicted in Figure 6, the goodness-of-fit indices are satisfactory (chi-square/*df* = 2.2, goodness-of-fit index = .914, and root mean square residual = .060). None of the other potential predictors of end-of-year performance was sufficiently related to final performance to merit additional comment.

Effects. One of the important outcomes of this kind of analysis is a consideration of the total effects of each structurally prior variable on our "ultimate" outcome measure, end-of-year science attainment. In Table 10 we present these total effects of all other variables. It can be seen that the indirect effect of home background (a combination of both mother's and father's educational levels and occupational prestige ratings) is .413, the effect of home support is .359, and the effect of entry level aptitude is 1.087. (That this last effect is larger than 1 suggests that our estimate of the reliability of the TOBE-2, .797, was a bit too low.) Also of interest in Table 10 is the pervasive effect of home background on all other latent variables except "percent of students on task during independent seatwork," an effect that was not included in these analyses.

[Insert Table 10 about here.]

First-Grade Analyses

As was the case for kindergarten, we began the first-grade analyses with a general model involving the home background indices, home-support activities, and students' beginning-of-year achievement levels; classroom observations of teacher behavior in terms of science instructional time, types of interactions and feedback, and praise and criticism; and end-of-year science performance. In addition, for first grade, we had results from the analysis of the text materials used in two of the three schools. (See Table 11 for a description of the variables originally specified for this analysis, as well as their final disposition.) Our initial model for the first-grade analysis is portrayed in Figure 7.

[Insert Table 11 about here.]

[Insert Figure 7 about here.]

Home Background Variables

The only noteworthy difference from kindergarten results for these variables at the first-grade level is that hours per week father worked no longer related to other variables in the model. Thus, the only variables remaining in the model are those representing parents' education and occupational prestige. Mother's education is again the variable with the largest loading (.86), and mother's occupation the smallest (.49). Through its effects on beginning-of-year performance (verbal and science) and the home support variables, home background has a total (indirect) effect on end-of-year science performance of .29.

Home Support

At the first-grade level, there were five variables in the home support category, in contrast to the three at the kindergarten level. We added questions about the number of school activities the students brought home and about their involvement in experiences that might specifically relate to science-relevant processes and activities. For this grade level, the index of literacy-related experiences was essentially unrelated to the other home support variables and was omitted from the final model. Of the four remaining variables, participation in activities had the largest loading (.64) on the composite index.

Beginning-of-Year Test Performance

Two composite indices were used at the first-grade level: verbal performance and science performance. The WRAT had the largest loading (.99) of the three tests assessing verbal performance, with both the Woodcock (.84) and Chicago (.77) contributing substantially. This composite indicator had a small and nonsignificant negative relationship to end-of-year science performance (β = -.08, t = -1.29).

Beginning science performance was measured by the two CIRCUS subtests: Listening and Think-It-Through. For these analyses, we constrained these two tests to have the same loading (estimated at .76) on the beginning-science composite. As expected, this beginning-of-year science measure was the latent variable that was by far the best predictor of end-of-year performance (β = .83).

Classroom Variables

As for the kindergarten analysis, we expected little by way of classroom contributions to end-of-year science performance. Based on the observational data, we estimated that first-grade students received only about 33 minutes per week of science instruction, so that there was little opportunity for teacher behavior to have any influence on acquisition of science knowledge. Also consistent with our experience in developing the kindergarten model was the finding that the measurement structure for variables in this category did not match our prior expectations. In contrast to the kindergarten results, where nearly all variables were combined into a single composite that we called "instructional style," there was no satisfactory, yet simple, structure relating these first-grade variables. The structure we finally accepted included two composite characteristics, labeled "instructional management" and "instructional style." Instruction management consisted of science instructional time (loading = .83) and procedural interactions (loading = .87), and instructional style was a combination of background knowledge interactions (loading = 1.00) and confirming feedback (loading = .78). In addition, we retained two classroom variables in their original form: sustaining feedback and text-tied interactions.

Of these four, only sustaining feedback had a significant relationship to end-of-year performance on the TOBE-2 (β = .15, t = 2.65). The instructional style and the text-tied interactions were both small and positive (β = .09 and .07, respectively), while the instructional management effect was small and negative (β = -.06).

Materials Analysis Variables

In this category are six variables (Table 11), five derived from the analysis of the instructional materials in the two schools that used any science textbooks and a sixth representing the number of domains from these materials that were actually taught while our observers were present. Only this sixth variable was retained in our final model, having a significant negative relationship to end-of-year performance (β = -.20, t = -2.98). That is, the more content domains from the textbook a teacher included in the science curriculum, the poorer the students' performance on the TOBE-2 at the end of the year.

[Insert Table 11 about here.]

End-of-Year Science Performance

We began this analysis with six indicators for end-of-year science performance: four scores from the Error Detection Test, the CIRCUS Think-It-Through score, and the total score on the TOBE-2. At this first-grade level, the Error Detection scores had only modest correlations with either the CIRCUS or the TOBE-2, and the CIRCUS and TOBE-2 were sufficiently different (r = .42) that it was not feasible to combine scores into some composite indicator of science performance. Since we then had to choose one of these test as our best indicator of performance, we selected the TOBE-2, both because its content was most representative of what we sought to measure and because it permitted us to maintain consistency with what we had done at the kindergarten level. We chose to use observed scores on the TOBE-2 as our index of performance without attempting to incorporate measurement error as we had at the kindergarten level. The consequence was to pool measurement error and prediction error into a single value representing the total proportion of unexplained variance in final performance. The resulting value, .334, indicated that we were able to "explain" about 2/3 of the variance in end-of-first-grade science performance by the other variables in the model.

Structural Model

Structural equation coefficients. Figure 8 depicts the final model for the influences on science performance at the end of first grade. Again, entry-level performance is the most important, with a structural coefficient of .825. Teachers' use of sustaining feedback, background knowledge interactions and confirming feedback (as a composite variable called instructional style), and text-tied interactions all make small positive contributions, while material usage observed and the combination of science instructional time and procedural interactions (management style) are both negatively related to final performance. Remember, though, that only the coefficients for entry-level performance, sustaining feedback, and material usage are statistically significant. "Fit" statistics for this model were as follows: chi-square/df = 2.2, goodness-of-fit index = .902, and root mean square residual = .061.

[Insert Figure 8 about here.]

Effects. The "effects" analysis for the first-grade model is presented in Table 12. The largest effect on end-of-year performance is, of course, that of beginning-of-year performance (.73), although it is lower than it was in kindergarten. The effect of home support (.42) is a bit larger than it was for kindergarten, while that for home background (.29) is correspondingly smaller. This same trend can be seen when comparing effects of these two composite variables throughout the two models. It may be that as these children proceed through the educational system, home background becomes less and less important while the kinds of ongoing home-based activities that might support school learning assume increasing importance.

[Insert Table 12 about here.]

Second-Grade Analyses

For the second-grade analyses, we had an initial pool of 48 observed variables, as described in Table 13, compared to 29 in kindergarten and 43 in first grade. Major additions at second grade were scores on the IRAS subtests and a new classroom interaction variable, interactions oriented toward fostering students' application of science concepts. Once again, we began our analyses with a variation on our general model in which all available data were considered. This initial model is represented in Figure 9.

[Insert Figure 9 about here.]

[Insert Table 13 about here.]

Home Background Variables

As was true for first grade, the parental education and occupational prestige measures formed a composite representing home background, but none of the other home background variables had any meaningful role in the model. As in the previous two analyses, mother's education was the primary variable represented in this composite (loading = .92), and mother's occupation the weakest (loading = .53). It is not surprising that these measures, obtained for all students in this study during their kindergarten year, have maintained a consistent structure over the three years with only minor variations from one year to another.

Home Support

Two of the variables that had been a part of this category for first grade were dropped from data collection for second grade, so that we had three measures of home behaviors supporting science learning: participation in activities, experiences with adults, and school activities brought home. Intercorrelations among these three are smaller than corresponding results in previous years (averaging .33 in kindergarten, .21 in first grade, and .14 in second grade). Once again, though, it was the participation in activities variable that had the highest loading (.62) on the composite index.

Beginning-of-Year Test Performance

At the first-grade level, we used two composite variables representing beginning-of-year capability. The first of these, verbal performance, combined the WRAT, Woodcock, five of the six IRAS subscores (dropping the self-corrects score, which had essentially no relationship to other measures), and--with a minimal loading--the Think-It-Through subtest from the CIRCUS battery. Four measures had comparably large loadings on the verbal composite: the WRAT, Woodcock, and word errors and comprehension questions scores from the IRAS, with loadings ranging from .87 to .92. The remaining IRAS subtests had loadings from .68 to .76, while the CIRCUS Think-It-Through subtest loaded only .09 on this verbal composite. Interestingly, beginning verbal performance had only a weak and insignificant relationship to final science performance (β = .02, t = .02).

The other composite, science performance, was formed by combining the two CIRCUS subtests (Listening and Think-It-Through), the four Error Detection subscores, and the TOBE-2 score from the end of first grade. CIRCUS-Listening, with a loading of .85, was the major observed indicator of this latent characteristic, while the remaining five measures had loadings ranging from .50 for sequence errors identified from the Error Detection Test to .66 for the Think-It-Through subtest from the CIRCUS. As in previous grades, this beginning science performance variable was by far the best predictor of end-of-year performance, with β = .88. In fact this is the strongest relationship of the three years of schooling.

Classroom Variables

At the second-grade level, although the amount of instructional time devoted to science had again increased from the previous year, the estimated 46 minutes a week average was still less than 10 minutes per day. Given such limited time, teachers' classroom behaviors were still not expected to have substantial influence on end-of-year science performance. Five variables from this category were retained in the final model, although only two of them--application interactions and amount of science time--had statistically significant relationships to end-of-year science performance, with coefficients of .183 and .175, respectively. The only composite variables emerging from this analysis were the simple summed values for praise (individual + class) and criticism (individual + class). The praise composite could easily be dropped from the model, having a non-0 relationship only to end-of-year verbal performance (β = -.06, t = -1.48), but criticism has a significant negative relationship to verbal performance (β = -.18) and a weaker negative relationship to science performance (β = -.13) at the end of second grade.

Materials Analysis Variables

As in first grade, only one variable in this category was retained in the final model: usage, or number of domains taught when an observer was present. The relationship of this variable was a modest one (β = -.12, t = -1.82), and it was once again negative. The more textbook domains a teacher included in the science curriculum, the poorer the students performed on the end-of-year measures of science understanding.

End-of-Year Performance

At the second grade level, we had two composite performance measures derived from the end-of-year tests: verbal, derived from the CIRCUS Reading (loading = .97) and STEP Science (loading = .49) subtests; and science, a composite of STEP Science (loading = .47) and the three locally developed tests, Plants (loading = .73), Three Forms of Matter (loading = .73), and Motion (loading = .33). The smaller loading for the Motion test is consistent with the fact that this topic essentially was not taught in any of these second-grade classrooms. It was intended to be an out-of-level test.

Structural Model

Structural equation coefficients. The model we have accepted as a reasonable, if not yet totally satisfactory, approximation to the structural relationships among these variables in second grade is presented in Figure 10. This model retains 32 of the original 48 variables, compared with 18 at the kindergarten level and 23 for first grade. For this model, the fit statistics were not quite as good as for the kindergarten and first-grade models: chi-square/*df* = 2.3, goodness-of-fit index = .836, and root mean square residual = .070.

Consistent with those previous models, beginning-of-year science performance has a substantially stronger relationship to final science performance than did any other variable (ß = .88). Noteworthy is the fact that entry-level verbal performance makes virtually no contribution (ß = .01) to final performance in science, once the other variables are included in the model. Application interactions and amount of science instructional time each with ß's of .18, are the other significant positive contributors to end-of-year science performance. Finally, observed material usage (ß = -.12) and teacher criticism (ß = -.13) are negatively related to final science performance, but not significantly so (*ts* = -1.83 and *t* = -1.38, respectively).

In this second-grade analysis, we also had two indicators of end-of-year verbal performance: the Reading subtest from the CIRCUS (loading = .97), and the STEP Science subtest (loading = .49). Although it is labeled a "science" measure, this STEP subtest has a substantial reading component to it. Not unexpectedly, beginning verbal performance was the best predictor of this composite (ß = .66), while beginning-of-year science performance also had a significant positive coefficient (ß = .30). Variables with a significant negative relationship to final verbal performance were teacher criticisms (ß = -.18) and science material usage (ß = -.10).

[Insert Figure 10 about here.]

Effects. A summary of effects for this analysis is presented in Table 14. Of course, the largest effect on end-of-year science performance is that of beginning-of-year science performance (1.02). Home background has an effect (.31) comparable to that from the first-grade analysis (.29), while the effect associated with the home support composite (.17) is considerably smaller than the (.42) obtained for first grade. It may be that the separation of verbal and science end-of-year performance accounts for this because the effect of home support on verbal performance (.38) is comparable in magnitude to its effect on science performance from the previous analysis. Whether the generally larger impact of home support activities on verbal than on science performance seen in Table 14 is peculiar to this sample or represents a dependable trend cannot be decided until additional analyses are carried out (e.g., for Cohort 2 at this grade level and for both cohorts at higher grades).

[Insert Table 14 about here.]

A Multi-Year Perspective

Based on the results of our attempts to use, across these early primary grades, indicators of (a) home environment, (b) student entry-level abilities, (c) home activities supportive of science learning, and d) classroom process indicators to account for end-of-year science achievement, we have arrived at several general conclusions regarding these results and the modeling process:

1. As we move up through the grades, the search for models becomes more challenging and less successful. As models have become more complex, we have been faced with more opportunities to make "wrong" decisions when seeking to refine our initial models for a grade level.

   In all structural modeling, a "final" model is just one of several alternative models that would fit the data equally well. When working in a domain in which theory is not strong enough to guide the modeler's decisions, as we have often been, any "final" model must be considered to be tentative. What we are reporting here, then, is a series of tentative models. What makes these efforts more than an academic exercise is the existence of (as yet unanalyzed) data from a second cohort, data that will be used for cross-validation of these derived models.

   We have also become increasingly concerned about the use of LISREL in these analyses because of the variance that we have noted between school districts. Future analyses of these data will examine the contexts of the individual districts.

2. A noteworthy trend that appears to be dependable in these models is the fading of the home support indices. In second grade, the three indices--experiences with adults, participation in activities, and school activities brought home--combined in a home support composite that had direct positive effects on both beginning-of-year verbal and science performance and, subsequently, indirect positive effects on end-of-year achievement. Yet the proportion of variance in end-of-year achievement that is accounted for in these models increases from grade to grade. Possible ways to account for this increase are considered next.

3. At second grade, there were four separate classroom variables influencing end-of-year science achievement: amount of science instructional time, materials usage, application interactions, and teacher criticisms. Their collective influence "explained" only about 6% of the variation in achievement. One interpretation of these various relationships is that teacher behaviors, as indicated by the classroom process indicators, had a greater effect in second grade than in earlier grades.

4. Therefore, we find that our hypothesis has been fairly well supported. Schooling did not affect students' science knowledge in kindergarten or first grade where so little time was spent on science instruction that one could hardly expect student ability to be altered by anything that happened there.

5. At the second-grade level, there was enough instructional time to make a difference in student performance, but apparently too few interactions that focused upon formulating hypotheses or questioning to mediate students' entering ability. It is also possible that at this grade level the pooled frequencies of these instructional characteristics mask substantial differences between districts. Future analyses will allow us to address this question.

Concluding Remarks

This report describes the conceptual basis and the data collection activities for a longitudinal study of how children learn science concepts over time. We have also presented basic descriptive statistics for those variables used in the "final" structural models we developed for grades K-2, along with diagrammatic representations and brief summary descriptions of the models themselves.

We believe that this study was a fair test of an attempt to determine what mediates children's science knowledge development in naturalistic settings that represent elementary schools in the United States at this time. We took entering ability, classroom processes, textbook characteristics, and children's outside activities into account. Therefore, we have been able to work with a data set that includes the potential major influences on children's lives when they are 5 to 8 years old. As pointed out earlier, we do see two trends. Home influences appear to decrease in influence on children's continuing science learning. Classroom processes, on the other hand, seem to mediate more differences in student achievement by second grade than they did in the two previous grades. These are hopeful findings, and they represent trends that would seem to be desirable because they suggest that schooling can be structured in ways that can make differences in students' achievement. It is also important to focus briefly on teaching practices that we have found to be negatively related to student achievement.

It is clear from our descriptive data, however, that there is very little science instruction taking place in the lower elementary grades. If these districts are representative of American schools today, so little time is given to science that if we had not been doing whole-day classroom observations we would no doubt have missed the little science instruction that did take place.

Furthermore, there was virtually no hands-on emphasis in science instruction we did observe. Teachers did not do demonstrations. Children had virtually no opportunity to manipulate materials, and there was very little use of science texts. Given this environment, it is not surprising that the United States is making a poor international showing in the upper grades.

It is promising that up to this point, the classroom variables we have found to make differences in children's performances in science are science instructional time, teachers' use of sustained feedback, and teachers' use of science concept application interactions. These findings have appeared elsewhere in the literature on general classroom processes for many years. Therefore, we are hopeful that many of the characteristics that have become hallmarks of successful teaching in other subject areas also appear to facilitate science learning.

It is very important, we believe, to emphasize that we have also found in both first and second grade that teachers' coverage of content has made consistent differences in student performance. Teachers who covered the largest number of content domains had *lower* student performance in science. Therefore, it appears that in-depth instruction in fewer areas enhances student learning in science. We hope that this will be received as good news by teachers. Our recommendation is that teachers select fewer areas of science to teach, then give children more instruction in those areas.

It is also important to examine these findings in the context of what we know to be the standing of the United States in comparison to foreign countries in science knowledge. We continue to rank far behind almost every other country studied by international surveys conducted during the last few years. While those comparisons typically begin with children in the middle grades or beyond, our results may shed some light on why we rank so low. We found so little science instruction in the earliest elementary grades that it would have been surprising to find that instruction had in fact mediated student performance to any great extent.

References

Abruscato, J., Fossaceca, J. W., Hassard, J., & Peck, D. (1986). *Holt elementary science*. New York: Holt, Rinehart & Winston.

Anderson, L. W. (1973). *Time and school learning*. Unpublished doctoral dissertation, University of Chicago.

Anderson, L. M., Evertson, C. M., & Brophy, J. E. (1978). An experimental study of effective teaching in first-grade reading groups. *Elementary School Journal, 79*, 193-223.

Anderson, T. H., & Armbruster, B. B. (1984). Content area textbooks. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools* (pp. 193-226). Hillsdale, NJ: Erlbaum.

Barr, R. (1983). *Chicago reading test*. Unpublished test, National College of Education, Evanston, IL.

Barr, R., & Dreeben, R., with Wiratchai, N. (1983). *How schools work*. Chicago: University of Chicago Press.

Block, H. H. (Ed.) (1971). *Mastery learning: Theory and practice*. New York: Holt.

Bredderman, T. (1983). Effects of activity-based science on student outcomes: A quantitative synthesis. *Review of Educational Research, 53*, 499-518.

Brophy, J. E., & Evertson, C. M. (1976) *Learning from teaching*. Boston: Allyn & Bacon.

Calfee, R. C., & Calfee, K. H. (1982). *Interactive reading assessment system (IRAS)*. Austin, TX: Southwest Educational Development Laboratory.

Carroll, J. (1963). A model for school learning. *Teachers College Record, 64*, 723-733.

Champagne, A. (1991). Misconceptions and teaching for conceptual change. In D. Alvermann & C. Santa (Eds.), *Science learning: Processes and application*. Newark, DE: International Reading Association.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Winfield, E. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: Office of Education, National Center for Educational Statistics, U.S. Government Printing Office.

Dahloff, U. S. (1971). *Ability grouping, content validity, and curriculum process analysis*. New York: Teachers College Press.

Educational Testing Service. (1976a). *CIRCUS reading test*. Menlo Park, CA: Addison-Wesley.

Educational Testing Service (1976b). *CIRCUS--Think it through*. Menlo Park, CA: Addison-Wesley.

Educational Testing Service. (1976c). *CIRCUS--Listen to the story*. Menlo Park, CA: Addison-Wesley.

Educational Testing Service. (1979). *Sequential test of educational progress*. Menlo Park, CA: Addison-Wesley.

Entwisle, D. R., & Hayduk, L. A. (1982). *Early schooling*. Baltimore, MD: Johns Hopkins Press.

Epstein, J. L. (1980). *A longitudinal study of school and family effects on student development.* Baltimore, MD: Johns Hopkins University, Center for Social Organization of Schools.

Fisher, C. W., Berliner, D. C., Filby, N. M., Marliave, R., Cahen, L. S., & Dishaw, M. M. (1980). Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), *Time to learn.* Washington, DC: USOE/NIE Printing.

Fort, D. C. (1990). From gifts to talents in science. *Phi Delta Kappan, 71*, 665-671.

Gilbert, S. W. (1989). An evaluation of the use of analogy, simile, and metaphor in science texts. *Journal of Research in Science Teaching, 26*(4), 315-328.

Good, T., & Grouws, D. (1979). The Missouri mathematics effectiveness project. *Journal of Educational Psychology, 71*, 355-362.

Hastings, C. N., Meyer, L. A., & Linn, R. L. (1987). *Assessing science concepts and process acquisition for first-, second-, and third-grade children.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Washington, DC.

Hastings, C. N., Meyer, L. A., Linn, R. L., & Wardrop, J. L. (in press). *Assessing science concepts and process acquisition for kindergarten, first-, second-, and third-grade children.* Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Heyns, B. (1978). *Summer learning and the effects of schooling.* New York: Academic Press.

Hofmeister, A., Engelmann, S., & Carnine, D. (1989). Developing and validating science education videodiscs. *Journal of Research in Science Teaching, 26*(8), 665-678.

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence.* New York: Basic Books.

Jastak, J. F., Jastak, S. R., & Bijou, S. W. (1978). *Wide range achievement test.* Wilmington, DE: Jastak Associates.

Jastak, S., & Wilkinson, G. S. (1984). *Wide range achievement test.* Wilmington, DE: Jastak Associates.

Joreskög, K. G. (1978). Structural analysis of correlation and covariance matrices. *Psychometrika, 43*, 443-447.

Joreskög, K. G., & Sorbom, D. (1984). *LISREL VI user's guide.* Mooresville, IN: Scientific Software.

Mallinson, G., G., Mallinson, J. B., Smallwood, W. L., & Valentino, C. (1985). *Silver Burdett science.* Morristown, NJ: Silver Burdett.

Mason, J., & Meyer, L. A. (1983). *The language and problem solving battery.* Unpublished test, University of Illinois, Center for the Study of Reading.

Meyer, L. A., Crummey, L., & Greer, E. A. (1988). Elementary science textbooks: Their contents, text characteristics, and comprehensibility. *Journal of Research in Science Teaching, 25*(6), 435-463.

Meyer, L. A., Greer, E. A., & Crummey, L. (1986). *Elementary science textbooks: Their contents, text characteristics and comprehensibility* (Tech. Rep. No. 386). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Greer, E. A., & Linn, R. L. (1985). *The error detection test*. Unpublished test, University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Greer, E. A., & Linn, R. L. (1986a). *Plants Test*. Unpublished test, University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., Greer, E. A., & Linn, R. L. (1986b). *Three forms of matter test*. Unpublished test, University of Illinois, Center for the Study of Reading.

Meyer, L. A., Hastings, C. N., & Linn, R. L. (1986). *Motion Test*. Unpublished test, University of Illinois, Center for the Study of Reading.

Moss, M. H. (1978a). *Test of basic experiences--2 (TOBE-2): Level K*. Monterey, CA: CTB/McGraw-Hill, Del Monte Research Park.

Moss, M. H. (1978b). *Test of basic experiences--2 (TOBE-2): Level L*. Monterey, CA: CTB/McGraw-Hill, Del Monte Research Park.

Piaget, J., & Inhelder, B. (1963). *The child's conception of space*. New York: Humanistic Press.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. M. W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand-McNally.

Sorenson, A. B., & Hallinan, M. T. (1977). A reconceptualization of school effects. *Sociology of Education, 50*, 273-289.

Stallings, J. A. (1976). *How instructional processes relate to child outcomes in a national study of follow through*. Menlo Park, CA: Stanford Research Institute.

Strube, P. (1989). The notion of style in physics textbooks. *Journal of Research in Science Teaching, 26*(4), 291-300.

Sund, R. B., Adams, D. K., & Hackett, J. K. (1982). *Accent on science*. Columbus, OH: Merrill.

Tifft, S. (1989, September 11). A crisis looms in science. *Time*, pp. 68-70.

Wainwright, C. L. (1989). The effectiveness of a computer-assisted instruction package in high school chemistry. *Journal of Research in Science Teaching, 26*(4), 275-290.

Woodcock, R. W. (1973). *Woodcock reading mastery tests*. Circle Pines, MN: American Guidance Service.

Author Note

We gratefully acknowledge the support and cooperation from the school district personnel and children in the three districts participating in this work. We are also indebted to our field staff: Mary Anderson, Joy Blair, Barbara Boyer, Eunice Buck, Lorraine Crummey, Marlene Engberg, Eunice Greer, Kathy Harper, Eleanor Hopke, Jennifer Hughes, Doris Krashow, Adina Lebowitz, Penny Lopez, Paul Mayberry, Linda Purcell, and Linda Ross, who completed classroom observations and the science textbook materials analysis.

Table 1

Means and Standard Deviations for all Measures of Student Ability

Instrument	Administered	$\bar{X}$	S.D.
WRAT Level 1 Reading Subtest	fall, kindergarten	19.07	7.28
Analogies Subtest of the Language and Problem Solving Battery	fall, kindergarten	5.45	2.77
Circus: Listen to the Story, Level A	fall, kindergarten	17.26	4.03
TOBE-2, Level K	spring, kindergarten	20.87	3.89
WRAT, Level 1 Reading Subtest	fall, first grade	28.54	7.94
Woodcock Passage Comprehension Test	fall, first grade	3.53	5.39
Chicago Reading Test	fall, first grade	31.51	24.60
Circus: Listening Level B	fall, first grade	30.36	5.49
Circus: Think It Through, Level B	fall, first grade	23.86	3.78
Error Detection: Word Errors identified	spring, first grade	8.82	1.58
Error Detection: Support for Word Errors	spring, first grade	8.34	2.04
Error Detection: Sequencing Errors Identified	spring, first grade	3.73	1.46
Error Detection: Support for Sequencing Errors	spring, first grade	3.05	1.53
TOBE-2, Level L	spring, first grade	21.71	2.83
WRAT Level 1 Reading Subtest	fall, second grade	49.26	9.84
Woodcock: Passage Comprehension Test	fall, second grade	21.53	10.66

Table 1 (Cont.)

Instrument	Administered	$\bar{X}$	S.D.
IRAS: Average Word List Errors	fall, second grade	0.37	0.96
IRAS: Average Word List Rate	fall, second grade	0.25	0.90
IRAS: Average Passage Errors	fall, second grade	0.26	0.89
IRAS: Average Passage Rate	fall, second grade	0.27	0.95
IRAS: Average Passage Self-Corrected Errors	fall, second grade	0.05	0.76
IRAS: Passage Comprehension Questions	fall, second grade	38.04	24.66
Circus: Think It Through, Level C	fall, second grade	24.71	5.69
Circus: Listening Level C	fall, second grade	33.25	5.17
Error Detection: Word Errors Identified	fall, second grade	8.99	1.18
Error Detection: Support for Word Errors	fall, second grade	8.57	1.55
Error Detection: Sequencing Errors Identified	fall, second grade	4.10	1.43
Error Detection: Support for Sequencing Errors	fall, second grade	3.11	1.59
Circus: Reading Test Level D	spring, second grade	31.53	7.83
STEP: Science	spring, second grade	36.11	7.38
Plants Test	spring, second grade	21.63	3.68
Three Forms of Matter Test	spring, second grade	20.31	5.23
Motion Test	spring, second grade	11.66	2.33

Table 2

Means and Standard Deviations by District for Time Spent in Kindergarten Science Instruction

District	N of Children	Time in Minutes/Observation $\bar{X}$	$\underline{SD}$	Percentage of Classes with 0 Minutes in Science
A	87	.87	.87	50
B	151	.52	.66	43
C	78	2.05	2.19	33

Table 3

Means and Standard Deviations by District for Interaction and Feedback Process Variables During Kindergarten Science Instruction

District	N	Interactions						Feedback					
		Procedural		Scriptal		Text-Tied		Sustaining		Terminal		Confirming	
		$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A	87	.89	.92	.45	.54	.67	.75	.23	.41	.05	.12	.21	.24
B	151	.08	.19	.24	.31	.22	.35	.00	.00	.00	.02	.04	.09
C	78	.16	.25	.82	1.13	.95	.97	.21	.31	.04	.07	.15	.15

Table 4

Means and Standard Deviations by District for Time Spent in First Grade Science Instruction

District	N of Children	Time in Minutes/Observation	
		$\bar{X}$	SD
A	88	6.62	4.09
B	152	5.68	2.66
C	79	8.22	4.39

Table 5

Means and Standard Deviations by District for Interaction and Feedback Process Variables During First Grade Science Instruction

		Interactions						Feedback					
		Procedural		Scriptal		Text-Tied		Sustaining		Terminal		Confirming	
District	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A	88	3.32	1.39	1.24	1.19	.49	.44	.12	.11	.10	.14	.92	.78
B	152	2.66	1.39	2.88	2.38	.46	.44	.38	.36	.12	.12	2.71	2.64
C	79	3.55	1.94	1.34	1.06	.59	.46	.13	.16	.03	.08	1.38	1.25

Table 6

Means and Standard Deviations by District for Time Spent in Second Grade Science Instruction

District	N of Children	Time in Minutes/Observation	
		$\bar{X}$	SD
A	78	5.32	4.04
B	144	10.40	4.09
C	89	12.53	8.37

Table 7

Means and Standard Deviations by District for Interaction and Feedback Process Variables During Second Grade Science Instruction

		Interactions						Feedback					
		Procedural		Scriptal		Text-Tied		Sustaining		Terminal		Confirming	
District	N	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
A	78	2.02	1.80	0.60	0.46	0.05	0.06	.01	.05	.03	.07	0.26	0.27
B	144	2.04	1.32	2.64	1.75	1.53	1.26	.18	.18	.09	.12	2.32	1.78
C	89	6.66	6.56	2.45	2.14	1.69	1.48	.53	.54	.19	.22	1.67	1.32

Table 8

Descriptive Statistics for Home Support Variables, All Grades and Schools

		School 1		School 2		School 3		Total Group	
Year	Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
K	PTCACTV	18.51	5.8	17.20	5.6	18.10	6.1	18.00	5.8
	EXPRWAD	31.89	5.8	30.88	5.2	32.04	6.2	31.55	5.6
	BKSMAGS	14.16	6.4	13.24	5.9	15.17	7.9	13.96	6.4
1	PTCACTV	10.93	5.9	9.54	6.0	9.94	5.9	10.32	5.9
	EXPRWAD	34.33	5.4	34.91	5.0	33.93	6.6	34.42	5.6
	SCHACTV	5.44	2.7	5.81	2.8	6.90	3.4	5.87	2.9
	PRACEXP	4.96	1.8	4.35	1.4	4.58	2.1	4.70	1.8
	LITEXPR	1.79	1.4	1.63	1.2	1.45	1.4	1.67	1.4
2	PTCACTV	10.65	6.2	11.08	5.8	10.73	6.9	10.78	6.3
	EXPRWAD	45.74	11.6	46.39	9.5	46.66	13.3	46.14	11.6
	SCHACTV	6.48	2.5	7.11	2.9	7.40	3.0	6.88	2.7
	SCHACTV	6.48	2.5	7.11	2.9	7.40	3.0	6.88	2.7

Table 9

Observed Indicators and Latent Variables: Kindergarten Analyses

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Home Background	Mother's Educ.	Home Bk. COMPOSITE	MTHREDUC
	Mother's Occup.	Home Bk. COMPOSITE	MTHROCC
	Father's Educ.	Home Bk. COMPOSITE	FATHEDUC
	Father's Occup.	Home Bk. COMPOSITE	FATHOCC
	Father: Hrs. per Week Working		FWKHRSWK
	Mother: Hrs. per Week Working		Dropped
	No. Yngr Siblings		Dropped
	No. Older Siblings		Dropped
	No. Adults in Home		Dropped
Home Activities Supporting Science Learning	Participation in Activities	Home Support COMPOSITE	PTCACTVK
	Experiences with Adults	Home Support COMPOSITE	EXPRWADK
	Literacy Related Experience	Home Support COMPOSITE	LITEXPRK
Beginning of Year Mental Tests	Analogies	Beg. Abil. COMPST.	ANLOGIES
	WRAT	Beg. Abil. COMPST.	WRATF83K
	Circus Listening	Beg. Abil. COMPST.	CRCSLSTK

Table 9 (Continued)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Classroom Variables	Science Instructional Time	Content Oriented Behavior (SUMMED)	SCITIMK
	Scriptal Interactions	Content Oriented Behavior (SUMMED)	MSCNACTK
	Text-Tied Interactions	Content Oriented Behavior (SUMMED)	MTXNACTK
	Content Oriented Behavior	Instructional Style COMPOSITE	CTORBEHK
	Procedural Interactions	Instructional Style COMPOSITE	MPRNACTK
	Sustaining Feedback	Instructional Style COMPOSITE	SUSFDBKK
	Confirming Feedback	Instructional Style COMPOSITE	CONFDBKK
	Terminating Feedback	Instructional Style COMPOSITE	TRMFDBKK
	Percent on Task During Seatwork		SWPCTTKK
	Frequency Seatwork Assigned		Dropped
	Teacher Criticisms: To Whole Class		Dropped
	Teacher Criticisms: To Indiv. Pupils		Dropped
	Teacher Praise: To Whole Class		Dropped
	Teacher Praise: To Indiv. Pupils		Dropped
End-of-Year Mental Test	TOBE-2	(Path fixed = .893)	TOBES84K

Table 10

Total Effects Represented by Kindergarten Structural Model

TOTAL EFFECTS OF KSI ON ETA

	FWKHRSWK	HOMBKGND	NSTRSTYL
ABILPREK	0.113	0.380	0.000
HOMESUPT	0.000	0.352	0.000
SWPCTTKK	0.000	0.000	0.442
SCPRFENK	0.123	0.413	-0.073

TOTAL EFFECTS OF KSI ON Y

	FWKHRSWK	HOMBKGND	NSTRSTYL
ANLOGIES	0.113	0.380	0.000
WRATF83K	0.102	0.342	0.000
CRCSLSTK	0.108	0.363	0.000
PTCACTVK	0.000	0.352	0.000
BKSMAGSK	0.000	0.311	0.000
EXPRWADK	0.000	0.284	0.000
SWPCTTKK	0.000	0.000	0.442
TOBES84K	0.110	0.369	-0.065

TOTAL EFFECTS OF ETA ON ETA

	ABILPREK	HOMESUPT	SWPCTTKK	SCPRFENK
ABILPREK	0.000	0.331	0.000	0.000
HOMESUPT	0.000	0.000	0.000	0.000
SWPCTTKK	0.000	0.000	0.000	0.000
SCPRFENK	1.087	0.359	0.000	0.000

TOTAL EFFECTS OF ETA ON Y

	ABILPREK	HOMESUPT	SWPCTTKK	SCPRFENK
ANLOGIES	1.000	0.331	0.000	0.000
WRATF83K	0.900	0.298	0.000	0.000
CRCSLSTK	0.954	0.315	0.000	0.000
PTCACTVK	0.000	1.000	0.000	0.000
BKSMAGSK	0.000	0.883	0.000	0.000
EXPRWADK	0.000	0.807	0.000	0.000
SWPCTTKK	0.000	0.000	1.000	0.000
TOBES84K	0.971	0.321	0.000	0.893

Table 11

Observed Indicators and Latent Variables: First Grade Analyses

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Home Background	Mother's Educ.	Home Bk. COMPOSITE	MTHREDUC
	Mother's Occup.	Home Bk. COMPOSITE	MTHROCC
	Father's Educ.	Home Bk. COMPOSITE	FATHEDUC
	Father's Occup.	Home Bk. COMPOSITE	FATHOCC
	Father: Hrs. per Week Working		Dropped
	Mother: Hrs. per Week Working		Dropped
	No. Yngr Siblings		Dropped
	No. Older Siblings		Dropped
	No. Adults in Home		Dropped
Home Activities Supporting Science Learning	Participation in Activities	Home Support COMPOSITE	PCTACTV1
	Experiences with Adults	Home Support COMPOSITE	EXPRWAD1
	Literacy Related Experience	Home Support COMPOSITE	Dropped
	School Activities Brought Home	Home Support COMPOSITE	SCHACTV1
	Process/Activity Experiences	Home Support COMPOSITE	PRACEXP1

Table 11 (Continued)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Beginning of Year Mental Tests	WRAT	Beg. Verbal Perf. COMPOSITE	WRATF841
	Woodcock	Beg. Verbal Perf. COMPOSITE	WDCKF841
	Chicago	Beg. Verbal Perf. COMPOSITE	CHGOF841
	Circus Listening	Beg. Science Perf. COMPOSITE	CRCSLST1
	Circus Think-It-Through	Beg. Science Perf. COMPOSITE	CRCSTTF1
Classroom Variables	Science Instructional Time	Instructional Mgmt. COMPOSITE	SCITIM1
	Procedural Interactions	Instructional Mgmt. COMPOSITE	MPRNACT1
	Scriptal Interactions	Instructional Style COMPOSITE	MSCNACT1
	Text-Tied Interactions		MTXNACT1
	Sustaining Feedback		SUSFDBK1
	Confirming Feedback	Instructional Style COMPOSITE	CONFDBK1
	Terminating Feedback		Dropped
	Percent on Task During Seatwork		Dropped
	Freq. Seatwork Assigned		Dropped
	Teacher Crit'sms: To Whole Class		Dropped

Table 11 (Continued)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Classroom Variables (Cont.)	Teacher Crit'sms: To Indiv. Pupils		Dropped
	Teacher Praise: To Whole Class		Dropped
	Teacher Praise: To Indiv. Pupils		Dropped
Materials Analysis	No. New Vocab. /Domain		Dropped
	No. Propositions /Domain		Dropped
	No. Questions /Domain		Dropped
	No. Activities /Domain		Dropped
	"Inconsid." Features/ Common Domain		Dropped
	No. Text Domains Observ. Taught	Standard Scores (Z)	MATUSG1Z
End-of-Year Mental Tests	Error Detection: Word Errors Id.		Dropped
	Err. Det.: Support for Wd. Errs.		Dropped
	Error Detection: Seq. Errors Id.		Dropped
	Err. Det.: Support for Seq. Errs.		Dropped
	TOBE-2		TOBES851

Table 12

Total Effects Represented by First Grade Structural Model

TOTAL EFFECTS OF KSI ON ETA

	HOMBKGND	NSTRMGMT	NSTRSTYL	SUSFDBK1	MATUSG1Z	MTXNACT1
VRPRFBG1	0.328	0.000	0.000	0.000	0.000	0.000
SCPRFBG1	0.426	0.000	0.000	0.000	0.000	0.000
HOMESUPT	0.304	0.000	0.000	0.000	0.000	0.000
SCPRFEN1	0.285	-0.055	0.082	0.127	-0.173	0.059

TOTAL EFFECTS OF KSI ON Y

	HOMBKGND	NSTRMGMT	NSTRSTYL	SUSFDBK1	MATUSG1Z	MTXNACT1
WRATF841	0.328	0.000	0.000	0.000	0.000	0.000
WDCKF841	0.279	0.000	0.000	0.000	0.000	0.000
CHGOF841	0.258	0.000	0.000	0.000	0.000	0.000
CRCSLST1	0.325	0.000	0.000	0.000	0.000	0.000
CRCSTTF1	0.325	0.000	0.000	0.000	0.000	0.000
PTCACTV1	0.304	0.000	0.000	0.000	0.000	0.000
SCHACTV1	0.175	0.000	0.000	0.000	0.000	0.000
EXPRWAD1	0.161	0.000	0.000	0.000	0.000	0.000
PRACEXP1	0.223	0.000	0.000	0.000	0.000	0.000
TOBES851	0.285	-0.055	0.082	0.127	-0.173	0.059

TOTAL EFFECTS OF ETA ON ETA

	VRPRFBG1	SCPRFBG1	HOMESUPT	SCPRFEN1
VRPRFBG1	0.000	0.000	0.274	0.000
SCPRFBG1	0.000	0.000	0.600	0.000
HOMESUPT	0.000	0.000	0.000	0.000
SCPRFEN1	-0.074	0.725	0.415	0.000

TOTAL EFFECTS OF ETA ON Y

	VRPRFBG1	SCPRFBG1	HOMESUPT	SCPRFEN1
WRATF841	1.000	0.000	0.274	0.000
WDCKF841	0.850	0.000	0.233	0.000
CHGOF841	0.785	0.000	0.215	0.000
CRCSLST1	0.000	0.763	0.458	0.000
CRCSTTF1	0.000	0.763	0.458	0.000
PTCACTV1	0.000	0.000	1.000	0.000
SCHACTV1	0.000	0.000	0.577	0.000
EXPRWAD1	0.000	0.000	0.529	0.000
PRACEXP1	0.000	0.000	0.733	0.000
TOBES851	-0.074	0.725	0.415	1.000

Table 13

Observed Indicators and Latent Variables: Second Grade Analyses

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Home Background	Mother's Educ.	Home Bk. COMPOSITE	MTHREDUC
	Mother's Occup.	Home Bk. COMPOSITE	MTHROCC
	Father's Educ.	Home Bk. COMPOSITE	FATHEDUC
	Father's Occup.	Home Bk. COMPOSITE	FATHOCC
	Father: Hrs. per Week Working		Dropped
	Mother: Hrs. per Week Working		Dropped
Home Activities Supporting Science Learning	Participation in Activities	Home Support COMPOSITE	PCTACTV2
	Experiences with Adults	Home Support COMPOSITE	EXPRWAD2
	School Activities Brought Home	Home Support COMPOSITE	SCHACTV2
Beginning-of-Year Mental Tests	WRAT	Beg. Verbal Perf. COMPOSITE	WRATF852
	Woodcock	Beg. Verbal Perf. COMPOSITE	WDCKF852
	IRAS: Rel. Errs. (Word Lists)	Beg. Verbal Perf. COMPOSITE	IRAREWF2
	IRAS: Rel. Rate (Word Lists)	Beg. Verbal Perf. COMPOSITE	IRARRWF2
	IRAS: Rel. Errs. (Passages)	Beg. Verbal Perf. COMPOSITE	IRAREPF2
	IRAS: Rel. Rate (Passages)	Beg. Verbal Perf. COMPOSITE	IRARRPF2

Table 13 (Cont.)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Beginning-of-Year Mental Tests (Cont.)	IRAS: Err. Self-Corrects (Psg)		Dropped
	IRAS: Comp. Qsts. (Passages)	Beg. Verbal Perf. COMPOSITE	IRAQSF2
	Circus Think-It-Through	Beg. Science Perf. (& Verbal Perf.)	CRCSTTF2
	Circus Listening	Beg. Science Perf. COMPOSITE	CRCSLSF2
	Error Detection: Word Errors Id.	Beg. Science Perf. COMPOSITE	EDWDERF2
	Err. Det.: Support for Wd. Errs.	Beg. Science Perf. COMPOSITE	EDWDSPF2
	Error Detection: Seq. Errors Id.	Beg. Science Perf. COMPOSITE	EDSQERF2
	Err. Det.: Support for Seq. Errs.	Beg. Science Perf. COMPOSITE	EDSQSPF2
	TOBE-2	Beg. Science Perf. COMPOSITE	TOBES851
Classroom Variables	Science Instructional Time		SCITIM2
	Procedural Interactions		MPRNACT2
	Application Interactions		MAPNACT2
	Scriptal Interactions		Dropped
	Text-Tied Interactions		Dropped
	Sustaining Feedback		Dropped

Table 13 (Cont.)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
Classroom Variables (Cont.)	Confirming Feedback		Dropped
	Terminating Feedback		Dropped
	Percent on Task During Seatwork		Dropped
	Freq. Seatwork Assigned		Dropped
	Teacher Crit'sms: To Whole Class	Criticism (SUMMED)	CRTCSM2
	Teacher Crit'sms: To Indiv. Pupils	Criticism (SUMMED)	CRTCSM2
	Teacher Praise: To Whole Class	Praise (SUMMED)	PRAISE2
	Teacher Praise: To Indiv. Pupils	Praise (SUMMED)	PRAISE2
Materials Analysis	No. New Vocab. /Domain		Dropped
	No. Propositions /Domain		Dropped
	No. Questions /Domain		Dropped
	No. Activities /Domain		Dropped
	No. Text Domains Observ. Taught	STANDARD SCORES (Z)	MATUSG2Z

Table 13 (Cont.)

General Category for Variable	Observed Variables	Transformations (if Used)	Acronym in Final Model (if Kept)
End-of-Year Mental Tests	Circus Reading	End Ver. Perf. COMPOSITE	CRCSRDS2
	STEP Science	End Ver. Perf. COMPOSITE End Sci. Perf. COMPOSITE	STEPSCS2
	Plants	End Sci. Perf. COMPOSITE	PLANTS2
	3 Forms of Matter	End Sci. Perf. COMPOSITE	F3MATRS2
	Motion	End Sci. Perf. COMPOSITE	MOTIONS2

Table 14

Total Effects Represented by Second Grade Structural Model

TOTAL EFFECTS OF KSI ON ETA

	HOMBKGND	SCITIME2	MATUSG2Z	MPRNACT2	MAPNACT2	PRAISE2	CRTCSM2
HOMESUPT	0.304	0.000	0.000	0.000	0.000	0.000	0.000
VRPRFBG2	0.221	0.000	0.000	0.000	0.000	0.000	0.000
SCPRFBG2	0.303	0.000	0.000	0.000	0.000	0.000	0.000
VRPRFEN2	0.294	0.086	-0.097	0.076	0.076	-0.060	-0.176
SCPRFEN2	0.311	0.127	-0.085	-0.044	0.133	0.000	-0.097

TOTAL EFFECTS OF KSI ON Y

	HOMBKGND	SCITIME2	MATUSG2Z	MPRNACT2	MAPNACT2	PRAISE2	CRTCSM2
EXPRWAD2	0.107	0.000	0.000	0.000	0.000	0.000	0.000
SCHACTV2	0.156	0.000	0.000	0.000	0.000	0.000	0.000
PTCACTV2	0.304	0.000	0.000	0.000	0.000	0.000	0.000
WRATF852	0.221	0.000	0.000	0.000	0.000	0.000	0.000
WDCKF852	0.213	0.000	0.000	0.000	0.000	0.000	0.000
CRCSLSF2	0.409	0.000	0.000	0.000	0.000	0.000	0.000
CRCSTTF2	0.342	0.000	0.000	0.000	0.000	0.000	0.000
IRAREWF2	-0.221	0.000	0.000	0.000	0.000	0.000	0.000
IRARRWF2	-0.183	0.000	0.000	0.000	0.000	0.000	0.000
IRAREPF2	-0.177	0.000	0.000	0.000	0.000	0.000	0.000
IRARRPF2	-0.163	0.000	0.000	0.000	0.000	0.000	0.000
IRQSF2	0.210	0.000	0.000	0.000	0.000	0.000	0.000
EDWDERF2	0.280	0.000	0.000	0.000	0.000	0.000	0.000
EDWDSPF2	0.315	0.000	0.000	0.000	0.000	0.000	0.000
EDSQERF2	0.241	0.000	0.000	0.000	0.000	0.000	0.000
EDSQSPF2	0.247	0.000	0.000	0.000	0.000	0.000	0.000
TOBS851	0.303	0.000	0.000	0.000	0.000	0.000	0.000
CRCSRDS2	0.294	0.000	0.000	0.000	0.000	0.000	0.000
STEPSCS2	0.349	0.086	-0.097	0.076	0.076	-0.060	-0.176
PLANTS2	0.311	0.126	-0.104	0.010	0.125	-0.030	-0.151
F3MATRS2	0.313	0.127	-0.085	-0.044	0.133	0.000	-0.097
MOTIONS2	0.140	0.128	-0.085	-0.044	0.134	0.000	-0.097
		0.057	-0.038	-0.020	0.060	0.000	-0.044

Table 14 (Continued)

TOTAL EFFECTS OF ETA ON ETA

	HOMESUPT	VRPRFBG2	SCPRFBG2	VRPRFEN2	SCPRFEN2
HOMESUPT	0.000	0.000	0.000	0.000	0.000
VRPRFBG2	0.432	0.000	0.000	0.000	0.000
SCPRFBG2	0.162	0.000	0.000	0.000	0.000
VRPRFEN2	0.377	0.701	0.459	0.000	0.000
SCPRFEN2	0.169	0.010	1.016	0.000	0.000

TOTAL EFFECTS OF ETA ON Y

	HOMESUPT	VRPRFBG2	SCPRFBG2	VRPRFEN2	SCPRFEN2
EXPRWAD2	0.352	0.000	0.000	0.000	0.000
SCHACTV2	0.512	0.000	0.000	0.000	0.000
PTCACTV2	1.000	0.000	0.000	0.000	0.000
WRATF852	0.432	1.000	0.000	0.000	0.000
WDCKF852	0.416	0.963	0.000	0.000	0.000
CRCSLSF2	0.218	0.000	1.349	0.000	0.000
CRCSTTF2	0.213	0.098	1.056	0.000	0.000
IRAREWF2	-0.433	-1.002	0.000	0.000	0.000
IRARRWF2	-0.358	-0.829	0.000	0.000	0.000
IRAREPF2	-0.346	-0.801	0.000	0.000	0.000
IRARRPF2	-0.319	-0.738	0.000	0.000	0.000
IRQSF2	0.410	0.951	0.000	0.000	0.000
EDWDERF2	0.149	0.000	0.000	0.000	0.000
EDWDSPF2	0.168	0.000	0.921	0.000	0.000
EDSQERF2	0.129	0.000	1.038	0.000	0.000
EDSQSPF2	0.132	0.000	0.794	0.000	0.000
TOBS851	0.162	0.000	0.812	0.000	0.000
CRCSRDS2	0.377	0.000	1.000	0.000	0.000
STEPSCS2	0.299	0.701	0.459	1.000	0.000
PLANTS2	0.169	0.360	0.890	0.504	0.648
F3MATRS2	0.170	0.010	1.016	0.000	1.000
MOTIONS2	0.076	0.010	1.024	0.000	1.007
		0.004	0.459	0.000	0.451

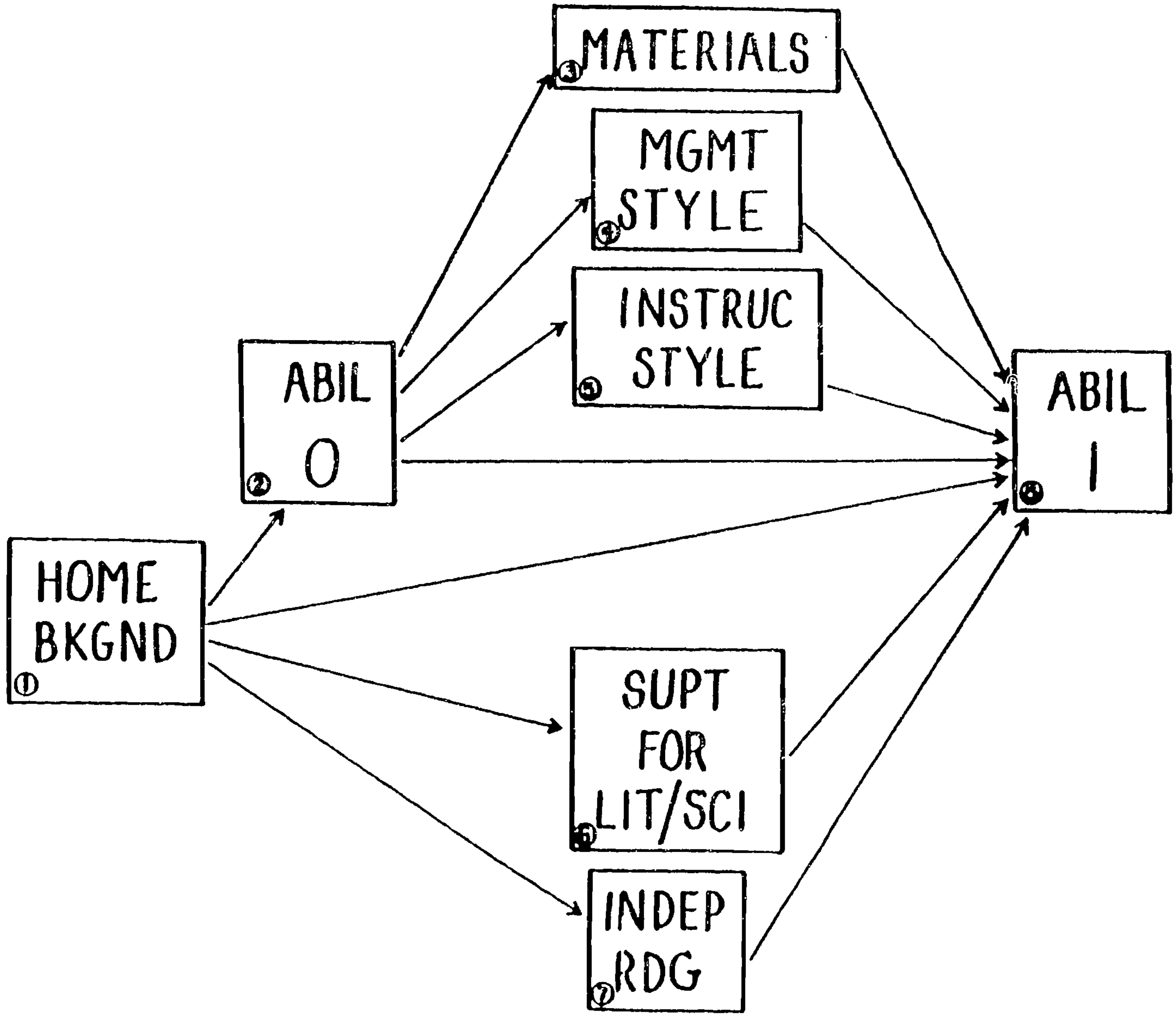


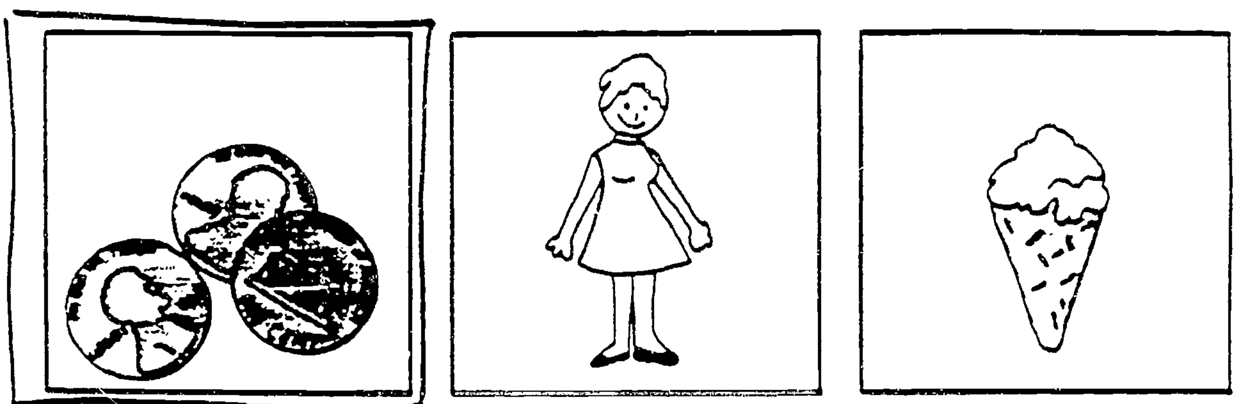
Figure 1
Heuristic Model of Science Knowledge Development

Figure 2
Comprehension Item
from the CIRCUS Listening Test

2

Look at these pictures and listen carefully. The teacher said. "Save your pennies so you can buy ice cream at the circus."

Mark what she told them to save.



Allow time. Then say:

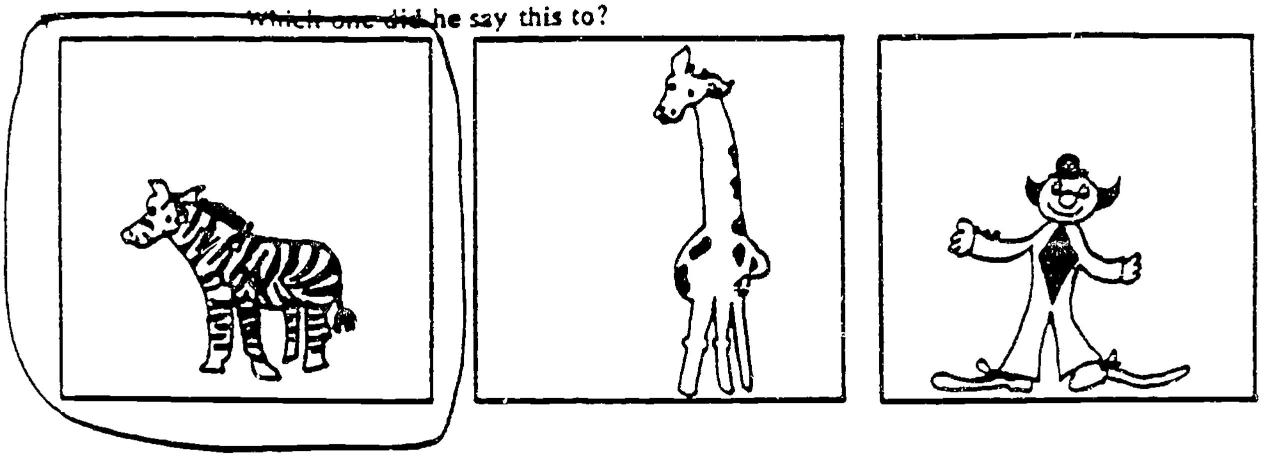
Now turn the page.

Figure 3
Interpretation Item
from the CIRCUS Listening Test

18

Clarence Clown wasn't always polite. The children heard him tell one of the animals to go take off his silly striped pajamas.

Which one did he say this to?



Turn the page.

Figure 4
Paragraph from the Kindergarten Science Questionnaire

Kindergarten
Science Questionnaire

As you read the next part of this questionnaire, put yourself into the story and answer as if you were telling this story about you and your kindergarten child. Circle the word(s) that describe you best. If another setting, such as a garden or garage is more typical for you and your child than the one we have picked, think of that setting as you choose your answers.

It is Saturday morning and you are about to start breakfast. You (strongly encourage, encourage, accept, discourage, strongly discourage) your children's presence in the kitchen. When your kindergartener is with you in the kitchen, you (very often, often, occasionally, seldom, never) talk to your child about what you are doing. For example, you are (very likely, likely, may, unlikely, very unlikely) to describe how to cut or fix things. In fact, your kindergartener has (very often, often, sometimes, seldom, never) fixed scrambled eggs and other things. Your child already seems to understand (very well, pretty well, well, poorly, very poorly) why he/she has to measure, mix, and cook most things before they are ready to eat. As your child helps in the kitchen, you (very often, often, sometimes, seldom, never) find yourself explaining how to do things, and why to do things. Your child can identify the uses of (10+, 7-9, 6-3, 2-0) gadgets and equipment in the kitchen. It is (very likely, somewhat likely, likely, somewhat unlikely, unlikely) that your kindergartener has had experiences helping an adult prepare a meal. Usually, this adult is (a man, a woman, both). And experiences like this occur in your house (10+, 6-9, 3-5, 1-2, 0) times a week.

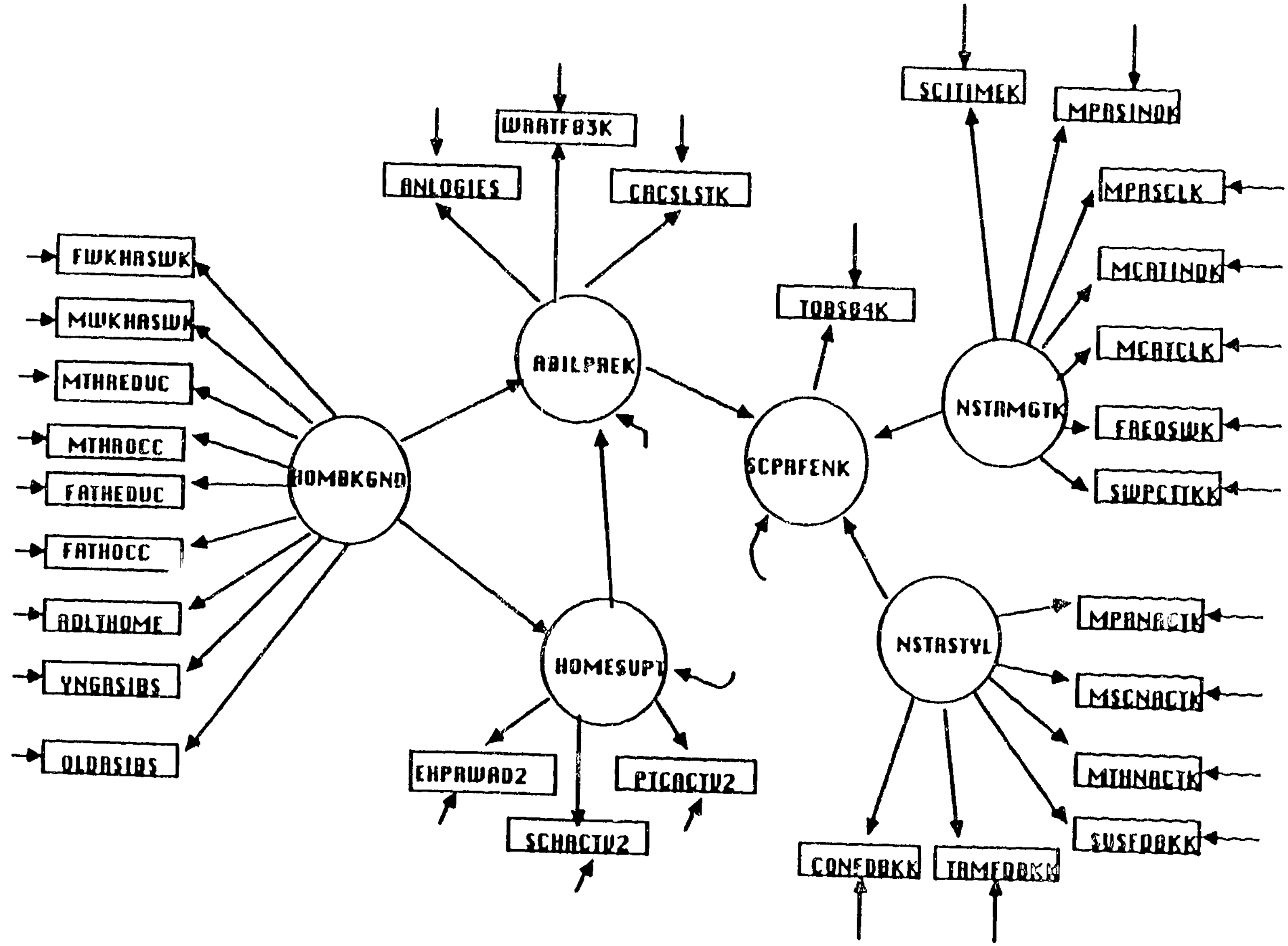


Figure 5
Initial LISREL Model for Kindergarten, Cohort 1

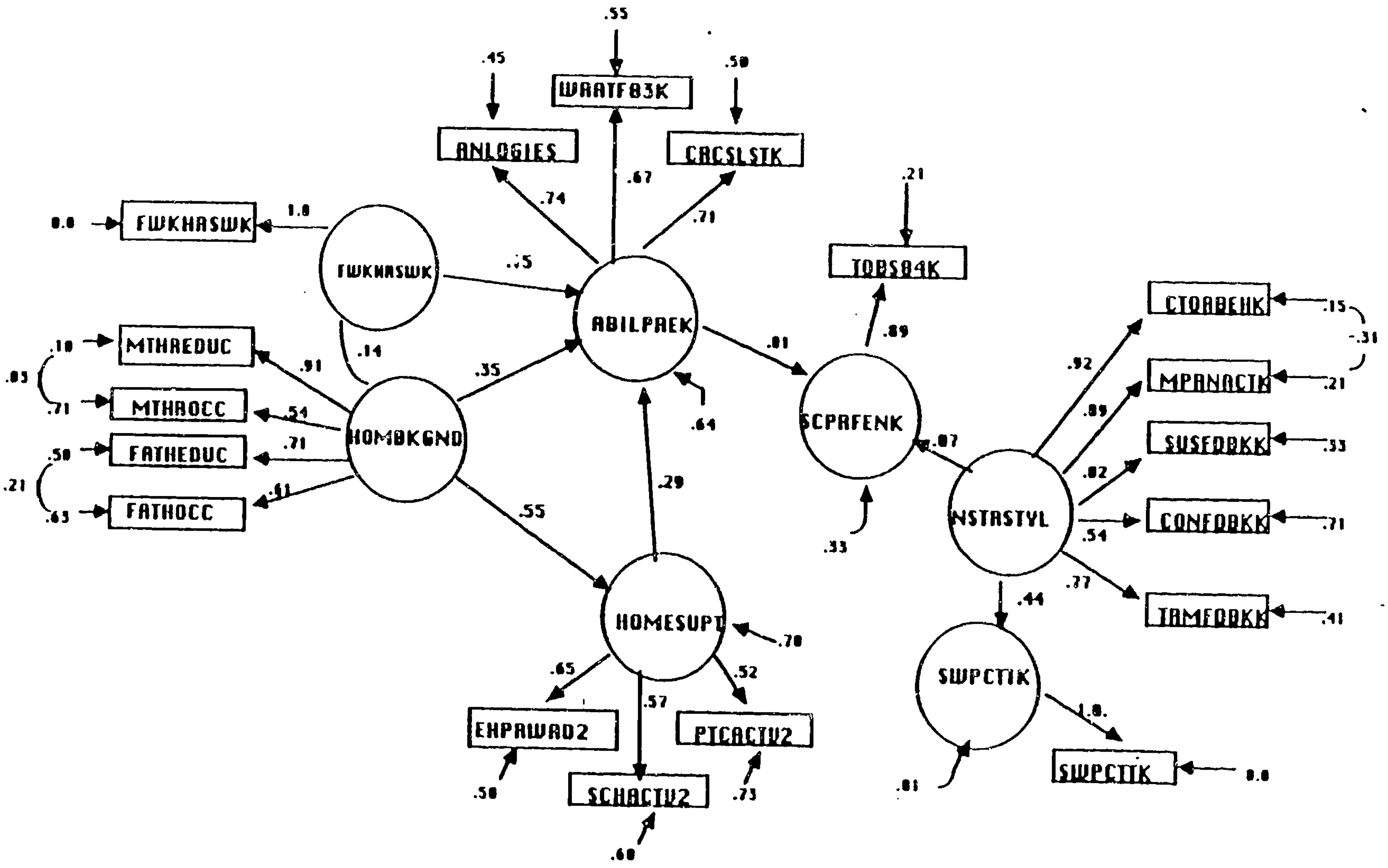


Figure 6
Final LISREL Model for Kindergarten, Cohort 1

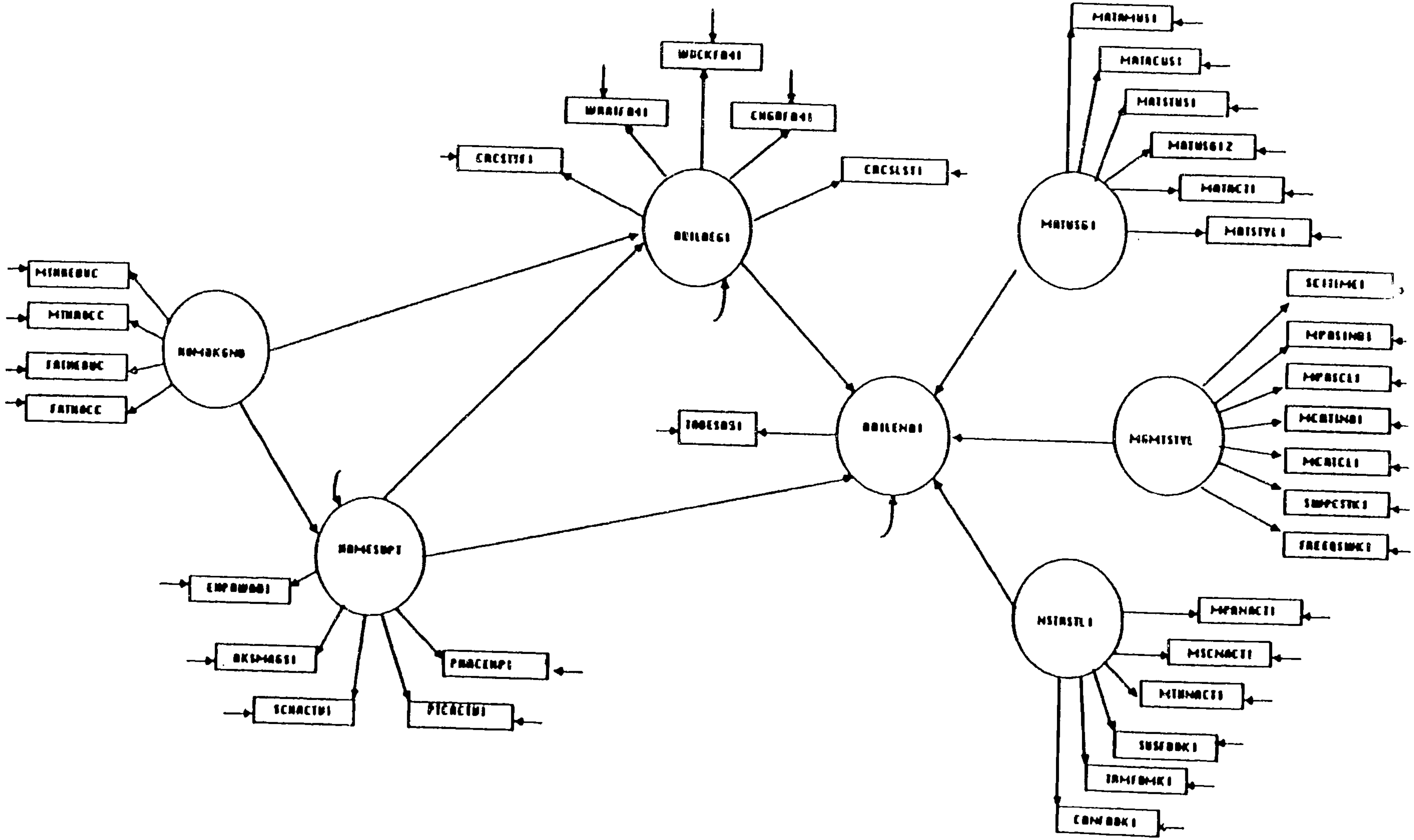


Figure 7
Initial LISREL Model for Grade 1, Cohort 1

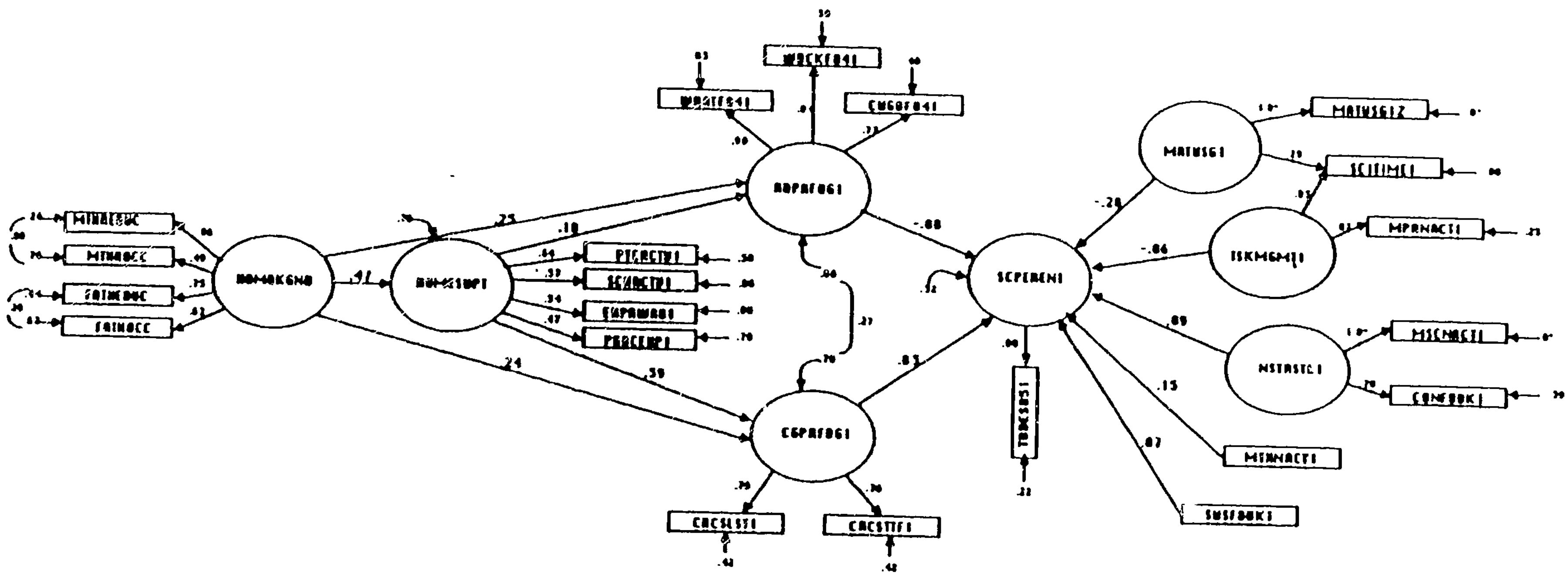


Figure 8
Final LISREL Model for Grade 1, Cohort 1

BEST COPY AVAILABLE

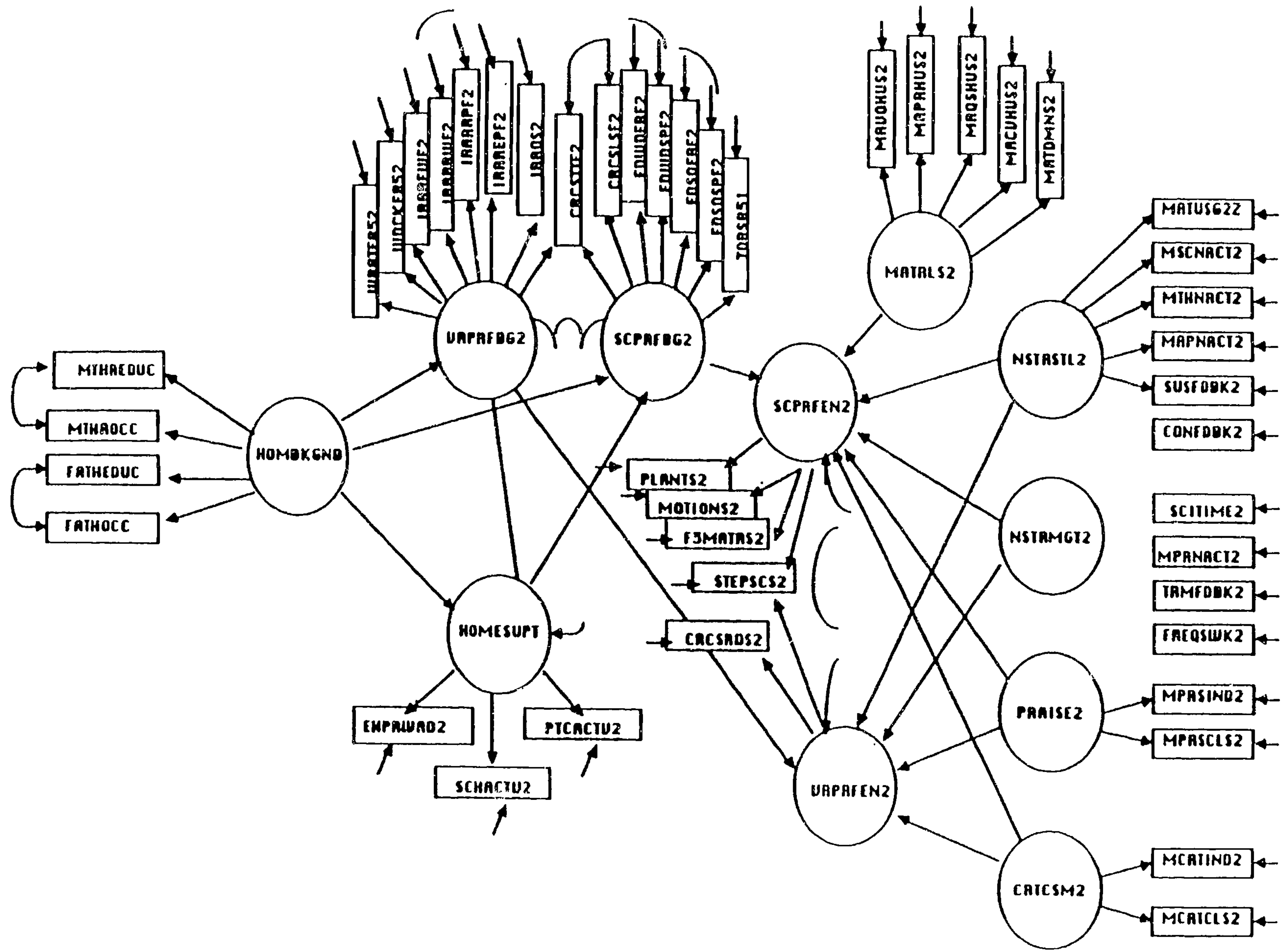


Figure 9
Initial LISREL Model for Grade 2, Cohort 1

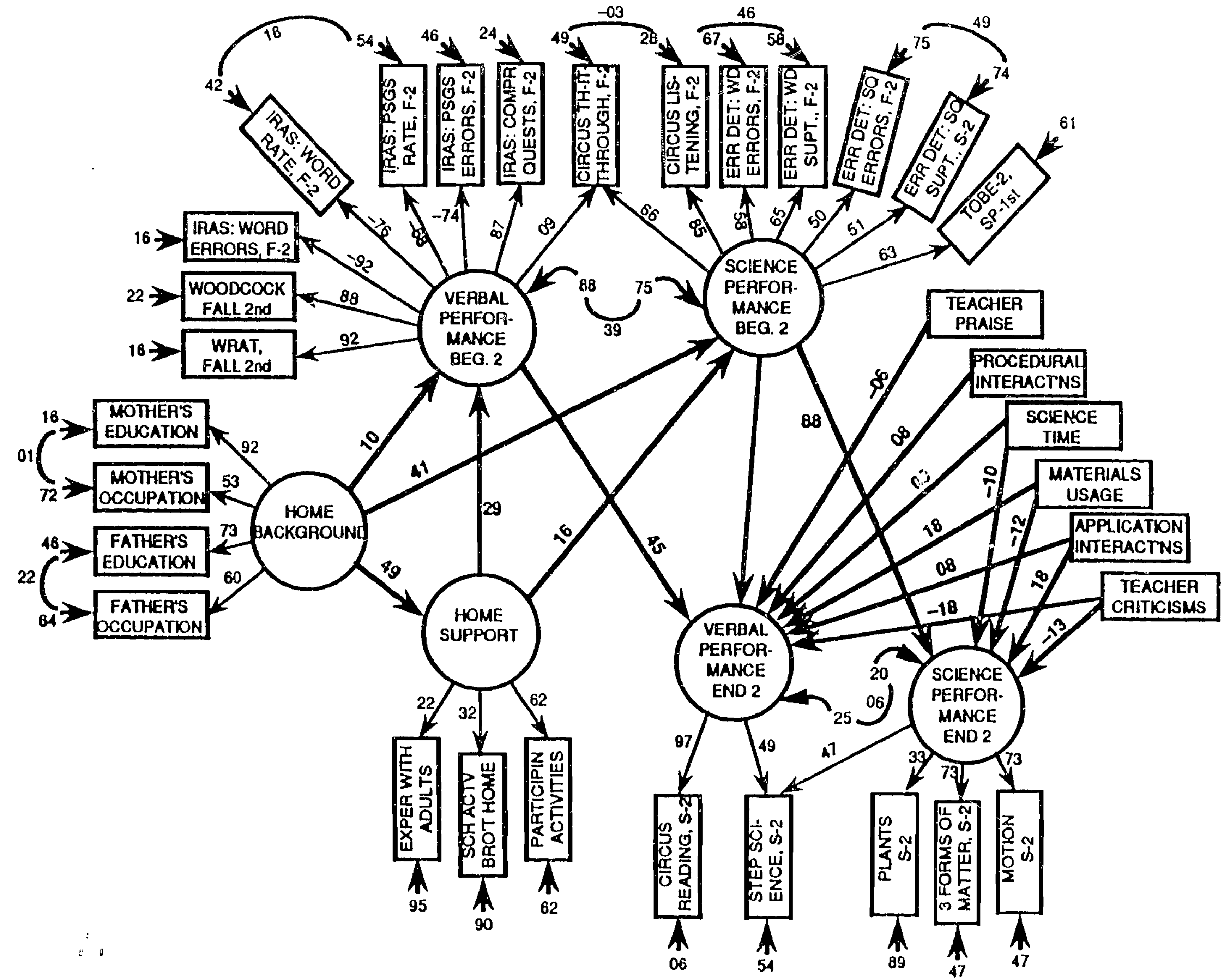


Figure 10
Final Structural Model for Science, Cohort 1, Grade 2